

Convex Optimization approach to signals with fast varying instantaneous frequency

Matthieu Kowalski^{a,b}, Adrien Meynard^a, Hau-tieng Wu^c

^a*Laboratoire des Signaux et Systèmes – Univ Paris-Sud – CNRS – CentraleSupélec*

^b*Parietal project-team, INRIA, Neurospin, CEA-Saclay, France*

^c*Department of Mathematics, University of Toronto, Toronto, Ontario, Canada*

Abstract

Motivated by the limitation of analyzing oscillatory signals composed of multiple components with fast-varying instantaneous frequency, we approach the time-frequency analysis problem by optimization. Based on the proposed adaptive harmonic model, the time-frequency representation of a signal is obtained by directly minimizing a functional, which involves few properties an “ideal time-frequency representation” should satisfy, for example, the signal reconstruction and concentrative time frequency representation. FISTA (Fast Iterative Shrinkage-Thresholding Algorithm) is applied to achieve an efficient numerical approximation of the functional. We coin the algorithm as *Time-frequency bY COncex OptimizationN* (Tycoon). The numerical results confirm the potential of the Tycoon algorithm.

Keywords: Time-frequency analysis, Convex optimization, FISTA, Instantaneous frequency, Chirp factor

1. Introduction

Extracting proper features from the collected dataset is the first step toward data analysis. Take an oscillatory signal as an example. We might ask how many oscillatory components inside the signal, how fast each component oscillates, how strong each component is, etc. Traditionally, Fourier transform is commonly applied to answer this question. However, it has been well known for a long time that when the signal is not composed of harmonic functions, then Fourier transform might not perform correctly. Specifically, when the signal satisfies $f(t) = \sum_{k=1}^K A_k(t) \cos(2\pi\phi_k(t))$, where $K \in \mathbb{N}$, $A_k(t) > 0$ and $\phi'_k(t) > 0$ but $A_k(t)$ and $\phi'_k(t)$ are not constants, the momentary behavior of the oscillation cannot be captured by the Fourier transform. A lot of efforts have been made in the past few decades to handle this problem. Time-frequency (TF) analysis based on different principals [21] has attracted a lot of attention in the field and

Email addresses: matthieu.kowalski@lss.supelec.fr (Matthieu Kowalski), hauwu@math.toronto.edu (Hau-tieng Wu)

Preprint submitted to Elsevier

March 22, 2016

many variations are available. Well known examples include short time Fourier transform (STFT), continuous wavelet transform (CWT), Wigner-Ville distribution (WVD), chirplet transform [39], S-transform [46], etc.

While these methods are widely applied in many fields, they are well known to be limited, again, by the Heisenberg uncertainty principle or the mode mixing problem caused by the interference known as the Moire patterns [21]. To alleviate the shortage of these analyses, in the past decades several solutions were proposed. For example, the empirical mode decomposition (EMD) [30] was proposed to study the dynamics hidden inside an oscillatory signal; however, its mathematical foundation is still lacking at this moment and several numerical issues cannot be ignored. Variations of EMD, like [51, 41, 24, 43, 20], were proposed to improve EMD. The sparsity approach [28, 26, 27, 47] and iterative convolution-filtering [36, 29, 12, 13] are another algorithms proposed to capture the flavor of the EMD, which have solid mathematical supports. The problem could also be discussed via other approaches, like the optimized window approach [44], nonstationary Gabor frame [3], ridge approach [44], the approximation theory approach [11], non-local mean approach [23] and time-varying autoregression and moving average approach [18], to name but a few. Among these approaches, the reassignment technique [33, 2, 8, 1] and the synchrosqueezing transform (SST) [16, 15, 9] have attracted more and more attention in the past few years. The main motivation of the reassignment technique is to improve the resolution issue introduced by the Heisenberg principal – the STFT coefficients are reallocated in *both* frequency axis and time axis according to their local phase information, which leads to the reassignment technique. The same reassignment idea can be applied to a very general settings like Cohen’s class, affine class, etc [22]. SST is a special reassignment technique; in SST, the STFT or CWT coefficients are reassigned *only* on the frequency axis [16, 15, 9] so that the causality is preserved and hence a real time algorithm is possible [10]. The same idea could be applied to different TF representation; for example, the SST based on wave packet transform or S-transform is recently considered in [52, 31].

By carefully examining these methods, we see that there are several requirements a time series analysis method for an oscillatory signal should satisfy. First, if the signal is composed of several oscillatory components with different frequencies, the method should be able to decompose them. Second, if the oscillatory component has time-varying frequency or amplitude, then how the frequency or amplitude change should be well approximated. Third, if any of the oscillatory component exists only over a finite period, the algorithm should provide a clear information about the starting point and ending point. Fourth, if we represent the oscillatory behavior in the TF plane, then the TF representation should be sharp enough and contain the necessary information. Fifth, the algorithm should be robust to noise. Sixth, the analysis should be adaptive to the signal we want to analyze. However, not every method could satisfy all these requirements. For example, due to the Heisenberg uncertainty principle, the TF representation of the STFT is blurred; the EMD is sensitive to noise and is incapable of handling the dynamics of the signal indicated in the third requirement. In addition to the above requirements, based on the problem we have interest, other features are needed from the TF analysis method, and some of them might not be easily fulfilled by the above approaches.

Among these methods, SST [16, 15, 9] and its variation [34, 52, 31, 42] could simultaneously satisfies these requirements, but it still has limitations. While SST could analyze oscillatory signals of “slowly varying instantaneous frequency (IF)” well with solid math-

ematical supports, the window needs to be carefully chosen if we want to analyze signals with fast varying IF [35]. Precisely, the conditions $|A'_k(t)| \leq \epsilon \phi'_k(t)$ and $|\phi''_k(t)| \leq \epsilon \phi'_k(t)$ are essential if we want to study the model $f(t) = \sum_{k=1}^K A_k(t) \cos(2\pi \phi_k(t))$ by the current SST algorithm proposed in [16, 15, 9]. Note that these “needs” could be understood/modeled as some suitable constraints, and to analyze the signal and simultaneously fulfill the designed constraints, optimization is a natural approach. Thus, in this paper, based on previous works and the above requirements, we would consider an optimization approach to study the oscillatory signals, which not only satisfies the above requirements, but also captures other features. In particular, we focus on capturing the fast varying IF. In brief, based on the relationship among the oscillatory components, the reconstruction property and the sparsity requirement on the time-frequency representation, we suggest to evaluate the optimal TF representation, denoted as F , by optimizing the following functional

$$\begin{aligned} \mathcal{H}(F, G) := & \int \left| \Re \int F(t, \omega) d\omega - f(t) \right|^2 dt \\ & + \mu \iint |\partial_t F(t, \omega) - i2\pi\omega F(t, \omega) + G(t, \omega) \partial_\omega F(t, \omega)|^2 dt d\omega \\ & + \lambda \|F\|_{L^1} + \gamma \|G\|_{L^2}, \end{aligned} \quad (1)$$

where G is an auxiliary function which quantifies the potentially fast varying instantaneous frequency. When G is fixed, it is clear that although $\mathcal{H}(\cdot, G)$ is not strictly convex, it is convex, so finding the minimizer is guaranteed. To solve this optimization problem, we propose to apply the widely applied and well studied algorithm *Fast Iterative Shrinkage-Thresholding Algorithm (FISTA)*. Embedded in an alternating minimization approach to estimate G and F , we coin the algorithm as *Time-frequency bY COnvex Optimization* (Tycoon).

The paper is organized in the following way. In Section 2, we discuss the adaptive harmonic model to model the signals with a fast varying instantaneous frequency and its identifiability problem; in Section 3, the motivation of the optimization approach based on the functional (1) is provided; in Section 4, we discuss the numerical details of Tycoon. In particular, how to apply the FISTA algorithm to solve the optimization problem; in Section 5, numerical results of Tycoon are provided.

2. Adaptive Harmonic Model

We start from introducing the model which we use to capture the signal with “fast varying IF”. The oscillatory signals with fast varying IF is commonly encountered in practice, for example, the chirp signal generated by bird’s song, bat’s vocalization and wolf’s howl, the uterine electromyogram signal, the heart rate time series of a subject with atrial fibrillation, the gravitational wave and the vibrato in violin play or human voice. More examples could be found in [22]. Thus, finding a way to study this kind of signal is fundamentally important in data analysis. First, we introduce the following model to capture the signals with fast varying IF, which generalizes the $\mathcal{A}_{\epsilon, d}^{c_1, c_2}$ class considered in [15, 9]:

Definition 2.1 (Generalized intrinsic mode type function (gIMT)). *Fix constants $0 \leq \epsilon \ll 1$, $c_2 > c_1 > \epsilon$ and $c_2 > c_3 > \epsilon$. Consider the functional set $\mathcal{Q}_\epsilon^{c_1, c_2, c_3}$, which consists of functions in $C^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$ with the following format:*

$$g(t) = A(t) \cos(2\pi\phi(t)), \quad (2)$$

which satisfies the following regularity conditions

$$A \in C^1(\mathbb{R}) \cap L^\infty(\mathbb{R}), \quad \phi \in C^3(\mathbb{R}), \quad (3)$$

the boundedness conditions for all $t \in \mathbb{R}$

$$\begin{aligned} \inf_{t \in \mathbb{R}} A(t) &\geq c_1, & \inf_{t \in \mathbb{R}} \phi'(t) &\geq c_1, \\ \sup_{t \in \mathbb{R}} A(t) &\leq c_2, & \sup_{t \in \mathbb{R}} \phi'(t) &\leq c_2, & \sup_{t \in \mathbb{R}} |\phi''(t)| &\leq c_3, \end{aligned} \quad (4)$$

and the growth conditions for all $t \in \mathbb{R}$

$$|A'(t)| \leq \epsilon \phi'(t), \quad |\phi'''(t)| \leq \epsilon \phi'(t). \quad (5)$$

Definition 2.2 (Adaptive harmonic model). *Fix constants $0 \leq \epsilon \ll 1$, $d > 0$ and $c_2 > c_1 > 0$. Consider the functional set $\mathcal{Q}_{\epsilon, d}^{c_1, c_2, c_3}$, which consists of functions in $C^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$ with the following format:*

$$g(t) = \sum_{\ell=1}^K g_\ell(t), \quad (6)$$

where K is finite and $g_\ell(t) = A_\ell(t) \cos(2\pi\phi_\ell(t)) \in \mathcal{Q}_\epsilon^{c_1, c_2, c_3}$; when $K > 1$, the following separation condition is satisfied:

$$\phi'_{\ell+1}(t) - \phi'_\ell(t) > d \quad (7)$$

for all $\ell = 1, \dots, K-1$.

We call ϵ, d, c_1, c_2 and c_3 *model parameters* of the $\mathcal{Q}_{\epsilon, d}^{c_1, c_2, c_3}$ model. Clearly, $\mathcal{Q}_{\epsilon, d}^{c_1, c_2, c_3} \subset \mathcal{Q}_\epsilon^{c_1, c_2, c_3}$ and both $\mathcal{Q}_\epsilon^{c_1, c_2, c_3}$ and $\mathcal{Q}_{\epsilon, d}^{c_1, c_2, c_3}$ are not vector spaces. Note that in the $\mathcal{A}_{\epsilon, d}^{c_1, c_2}$ model, the condition “ $\phi_\ell \in C^3(\mathbb{R})$, $\sup_{t \in \mathbb{R}} |\phi''_\ell(t)| \leq c_2$ and $|\phi'''_\ell(t)| \leq \epsilon \phi'_\ell(t)$ for all $t \in \mathbb{R}$ ” is replaced by “ $\phi_\ell \in C^2(\mathbb{R})$ and $|\phi''_\ell(t)| \leq \epsilon \phi'_\ell(t)$ for all $t \in \mathbb{R}$ ”. Thus, we say that the signals in $\mathcal{A}_{\epsilon, d}^{c_1, c_2}$ are oscillatory with slowly varying instantaneous frequency. Also note that $\mathcal{A}_{\epsilon, d}^{c_1, c_2}$ is not a subset of $\mathcal{Q}_{\epsilon, d}^{c_1, c_2, c_3}$. Indeed, for $A_\ell(t) \cos(2\pi\phi_\ell(t)) \in \mathcal{A}_{\epsilon, d}^{c_1, c_2}$, even if $\phi_\ell \in C^3(\mathbb{R})$, the third order derivative of ϕ_ℓ is not controlled. Also note that the number of possible components K is controlled by the model parameters; that is, $K \leq \frac{c_2 - c_1}{d}$.

Remark. We have some remarks about the model. First, note that it is possible to introduce more constants to control $A(t)$, like $0 < c_4 \leq \inf_{t \in \mathbb{R}} A(t) \leq \sup_{t \in \mathbb{R}} A(t) \leq c_5$, in addition to the control of ϕ' by $c_1, c_2 > 0$ in the model. Also, to capture the “dynamics”, we could consider a more general model dealing with the “sudden appearance/disappearance”, like $g(t) = \sum_{\ell=1}^K g_\ell(t) \chi_{I_\ell}$, where χ is the indicator function and $I_\ell \subset \mathbb{R}$ is connected and long enough. However, while these will not generate fundamental differences but will complicate the notation, to simplify the discussion, we stick to our current model.

Second, we could consider different models to study the “fast varying IF”. For example, we could replace the condition “ $|A'(t)| \leq \epsilon \phi'(t)$, $\phi_\ell \in C^3(\mathbb{R})$, $\sup_{t \in \mathbb{R}} |\phi_\ell''(t)| \leq c_2$ and $|\phi_\ell'''(t)| \leq \epsilon \phi_\ell'(t)$ for all $t \in \mathbb{R}$ ” by the slow evolution chirp conditions [22]; that is “ $|A'(t)| \leq \epsilon A(t) \phi'(t)$, $\phi_\ell \in C^2(\mathbb{R})$ and $|\phi_\ell''(t)| \leq \epsilon \phi_\ell'(t)^2$ for all $t \in \mathbb{R}$ ”. We refer the reader with interest in the detailed discussion about this “slow evolution chirp model” to [22, Section 2.2]. A simplified slow evolution chirp model (with the condition $|A'(t)| \leq \epsilon \phi'(t)$) is recently considered in [37] for the study of the sparsity approach to TF analysis. We mention that the argument about the identifiability issue stated below for $\mathcal{Q}_{\epsilon, d}^{c_1, c_2, c_3}$ could be directly applied to state the identifiability issue of the slow evolution chirp model.

Before proceeding to say what it means by “instantaneous frequency” or “amplitude modulation”, we immediately encounter a problem which is understood as the *identifiability problem*. Indeed, we might have infinitely many different ways to represent a cosine function $g_0(t) = \cos(2\pi t)$ in the format $a(t) \cos(2\pi \phi(t))$ so that $a > 0$ and $\phi' > 0$, even though it is well known that $g_0(t)$ is a harmonic function with amplitude 1 and frequency 1. Precisely, there exist infinitely many smooth functions α and β so that $g_0(t) = \cos(2\pi t) = (1 + \alpha(t)) \cos(2\pi(t + \beta(t)))$, and in general there is no reason to favor $\alpha(t) = \beta(t) = 0$. Before resolving this issue, we could not take amplitude 1 and frequency 1 as reliable features to quantify the signal g_0 when we view it as a component in $\mathcal{Q}_{\epsilon}^{c_1, c_2, c_3}$. In [9], it is shown that if $g(t) = A(t) \cos(2\pi \phi(t)) = [A(t) + \alpha(t)] \cos(2\pi[\phi(t) + \beta(t)])$ are both in $\mathcal{A}_{\epsilon, d}^{c_1, c_2}$, then $|\alpha(t)| \leq C\epsilon$ and $|\beta'(t)| \leq C\epsilon$, where C is a constant depending only on the model parameters c_1, c_2, d . Therefore, A_ℓ and ϕ'_ℓ are unique locally up to an error of order ϵ , and hence we could view them as features of an oscillatory signal in $\mathcal{A}_{\epsilon, d}^{c_1, c_2}$. Here, we show a parallel theorem describing the identifiability property for the functions in the $\mathcal{Q}_{\epsilon, d}^{c_1, c_2, c_3}$ model.

Theorem 2.1 (Identifiability of $\mathcal{Q}_{\epsilon}^{c_1, c_2, c_3}$). *Suppose a gIMT $a(t) \cos \phi(t) \in \mathcal{Q}_{\epsilon}^{c_1, c_2, c_3}$ can be represented in a different form which is also a gIMT in $\mathcal{Q}_{\epsilon}^{c_1, c_2, c_3}$; that is, $a(t) \cos \phi(t) = A(t) \cos \varphi(t) \in \mathcal{Q}_{\epsilon}^{c_1, c_2, c_3}$. Define $t_m := \phi^{-1}((m + 1/2)\pi)$ and $s_m := \phi^{-1}(m\pi)$, $m \in \mathbb{Z}$, $\alpha(t) := A(t) - a(t)$, and $\beta(t) := \varphi(t) - \phi(t)$. Then we have the following controls of α and β at t_m and s_m*

1. Up to a global factor $2l\pi$, $l \in \mathbb{Z}$, $\beta(t_n) = 0$ for all $n \in \mathbb{Z}$;
2. $\frac{a(t_n)}{a(t_n) + \alpha(t_n)} = \frac{\phi'(t_n) + \beta'(t_n)}{\phi'(t_n)}$ for all $n \in \mathbb{Z}$. In particular, $\alpha(t_n) = 0$ if and only if $\beta'(t_n) = 0$ for all $n \in \mathbb{Z}$;
3. $\frac{a(s_n)}{a(s_n) + \alpha(s_n)} = \cos(\beta(s_n))$ for all $n \in \mathbb{Z}$. In particular, $\alpha(s_m) = 0$ if and only if $\beta(s_m) = 0$, $m \in \mathbb{Z}$.

Furthermore, the size of α and β are bounded by

1. $|\alpha(t)| < 2\pi\epsilon$ for all $t \in \mathbb{R}$;
2. $|\beta''(t)| \leq 2\pi\epsilon$, $|\beta'(t)| \leq \frac{2\pi\epsilon}{c_1}$ and $|\beta(t)| \leq \frac{2\pi\epsilon}{c_1^2}$ up to a global factor $2l\pi$, $l \in \mathbb{Z}$, for all $t \in \mathbb{R}$.

We mention that the controls of α and β at t_m and s_m do not depend on the growth condition in (5). However, to control the size of α and β , we need the growth condition in (5).

Theorem 2.2 (Identifiability of $\mathcal{Q}_{\epsilon,d}^{c_1,c_2,c_3}$). Suppose $f(t) \in \mathcal{Q}_{\epsilon,d}^{c_1,c_2,c_3}$ can be represented in a different form which is also in $\mathcal{Q}_{\epsilon,d}^{c_1,c_2,c_3}$; that is,

$$f(t) = \sum_{l=1}^N a_l(t) \cos \phi_l(t) = \sum_{l=1}^M A_l(t) \cos \varphi_l(t) \in \mathcal{Q}_{\epsilon,d}^{c_1,c_2,c_3}. \quad (8)$$

Then, when $d \geq \sqrt{2 \ln c_2 + \frac{1}{2} \ln c_3 - \ln \epsilon}$, $M = N$ and for all $t \in \mathbb{R}$ and for all $l = 1, \dots, N$, the following holds:

1. $|\phi_l(t) - \varphi_l(t)| = O(\sqrt{\epsilon})$ up to a global factor $2n\pi$, $n \in \mathbb{Z}$;
2. $|\phi'_l(t) - \varphi'_l(t)| = O(\sqrt{\epsilon})$;
3. $|\phi''_l(t) - \varphi''_l(t)| = O(\sqrt{\epsilon})$;
4. $|a_l(t) - A_l(t)| = O(\sqrt{\epsilon})$,

where the constants on the right hand side are universal constants depending on the model parameters of $\mathcal{Q}_{\epsilon,d}^{c_1,c_2,c_3}$.

Note that in this theorem, the bound $\sqrt{\epsilon}$ and the lower bound of d are by no means optimal since we consider the case when there are as many components as possible. We focus on showing that even when there are different representations of a given function in $\mathcal{Q}_{\epsilon,d}^{c_1,c_2,c_3}$, the quantities we have interest are close up to a negligible constant. As a result, we have the following definitions, which generalize the notion of amplitude and frequency.

Definition 2.3. [Phase function, instantaneous frequency, chirp factor and amplitude modulation] Take a function $f(t) = \sum_{\ell=1}^N a_\ell(t) \cos \phi_\ell(t) \in \mathcal{Q}_{\epsilon,d}^{c_1,c_2,c_3}$. For each $\ell = 1, \dots, N$, the monotonically increasing function $\phi_\ell(t)$ is called the phase function of the ℓ -th gIMT; the first derivative of the phase function, $\phi'_\ell(t)$, is called the instantaneous frequency (IF) of the ℓ -th gIMT; the second derivative of the phase function, $\phi''_\ell(t)$, is called the chirp factor (CF) of the ℓ -th gIMT; the positive function $A_\ell(t)$ is called the amplitude modulation (AM) of the ℓ -th gIMT.

Note that the IF and AM are always positive, but usually not constant. On the other hand, the CF might be negative and non-constant. Clearly, when ϕ_ℓ are all linear functions with positive slopes and A_ℓ are all positive constants, then the model is reduced to the harmonic model and the IF is equivalent to the notion frequency in the ordinary Fourier transform sense. The conditions $|A'_\ell(t)| \leq \epsilon \phi'_\ell(t)$ and $|\phi'''_\ell(t)| \leq \epsilon \phi''_\ell(t)$ force the signal to locally behave like a harmonic function or a chirp function, and hence the nominations. By Theorem 2.1 and Theorem 2.2, we know that the definition of these quantities are unique up to an error of order ϵ .

We could also model the commonly encountered ingredient in signal processing – the shape function, trend and noise as those considered in [50, 9]. However, to concentrate the discussion on the optimization approach to the problem, in this paper we focus only on the $\mathcal{Q}_{\epsilon,d}^{c_1,c_2,c_3}$ functional class.

3. Optimization Approach

In general, given a function $f(t) = \sum_{k=1}^K A_k(t) \cos(2\pi\phi_k(t))$ so that $A_k(t) > 0$ and $\phi'_k(t) > 0$ for $t \in \mathbb{R}$, we would expect to have the *ideal time-frequency representation*

(iTFR), denoted as $R_f(t, \omega)$, satisfying

$$R_f(t, \omega) = \sum_{k=1}^K A_k(t) e^{i2\pi\phi_k(t)} \delta_{\phi'_k(t)}(\omega), \quad (9)$$

where $\delta_{\phi'_k(t)}$ is the Dirac measure supported at $\phi'_k(t)$, so that we could well extract the features $A_k(t)$ and $\phi'_k(t)$ describing the oscillatory signal from R_f . Note that the iTFR is a distribution. In addition, the reconstruction and visualization of each component are possible. Indeed, we can reconstruct the k -th component by integrating along the frequency axis on the period near $\phi'_k(t)$. Indeed,

$$A_k(t) \cos(2\pi\phi_k(t)) = \Re \int_{\mathbb{R}} R_f(t, \omega) \psi \left(\frac{\omega - \phi'_k(t)}{\theta} \right) d\omega, \quad (10)$$

where \Re means taking the real part, $\theta \ll 1$, ψ is a compactly supported Schwartz function so that $\psi(0) = 1$. Further, the visualization is realized via displaying the “time-varying power spectrum” of f , which is defined as

$$S_f(t, \omega) := \sum_{k=1}^K A_k^2(t) \delta_{\phi'_k(t)}(\omega), \quad (11)$$

and we call it the *ideal time-varying power spectrum* (itvPS) of f , which is again a distribution.

To evaluate the iTFR for a function $f = \sum_{k=1}^K A_k(t) \cos(2\pi\phi_k(t))$, we fix $0 < \theta \ll 1$ and consider the following *approximative iTFR with resolution θ*

$$\tilde{R}_f(t, \omega) = \sum_{k=1}^K A_k(t) e^{i2\pi\phi_k(t)} \frac{1}{\theta} h \left(\frac{\omega - \phi'_k(t)}{\theta} \right), \quad (12)$$

where $t \in \mathbb{R}$, $\omega \in \mathbb{R}$ and h is a Schwartz function supported on $[-\sigma, \sigma]$, $\sigma > 0$, so that $\int h = 1$ and $\frac{1}{\epsilon} h(\frac{\cdot}{\epsilon})$ converges to Dirac measure δ supported at 0 weakly as $\epsilon \rightarrow 0$ and $\int h(x) dx = 1$. Clearly, we know that \tilde{R}_f is essentially supported around $(t, \phi'_k(t))$ for $k = 1, \dots, K$ and as $\theta \rightarrow 0$, \tilde{R}_f converges to the iTFR in the weak sense. Also, we have for all $t \in \mathbb{R}$ and $k = 1, \dots, K$, when θ is small enough so that $\sigma\theta > d$ is satisfied, where d is the constant defined in the separation condition in (7), we have

$$\Re \int_{\phi'_k(t) - \sigma\theta}^{\phi'_k(t) + \sigma\theta} \tilde{R}_f(t, \omega) d\omega = A_k(t) \cos(2\pi\phi_k(t)). \quad (13)$$

Thus, the reconstruction property of iTFR is satisfied. In addition, the visualization property of itvPS can be achieved by taking

$$\tilde{S}_f(t, \omega) = \left| \tilde{R}_f(t, \omega) \right|^2 = \sum_{k=1}^K |A_k(t)|^2 \frac{1}{\theta^2} \left| h \left(\frac{\omega - \phi'_k(t)}{\theta} \right) \right|^2, \quad (14)$$

where the equality holds due to the facts that ϕ'_k are separated and $\theta \ll 1$. Next we need to find other conditions about \tilde{R}_f . A natural one is observing its differentiation. By a

direct calculation, we know $\frac{1}{\theta^2} h' \left(\frac{\omega - \phi'_k(t)}{\theta} \right) = \partial_\omega \frac{1}{\theta} h \left(\frac{\omega - \phi'_k(t)}{\theta} \right)$, and hence we have

$$\begin{aligned}
\partial_t \tilde{R}_f(t, \omega) &= \sum_{k=1}^K A'_k(t) e^{i2\pi\phi(t)} \frac{1}{\theta} h \left(\frac{\omega - \phi'_k(t)}{\theta} \right) \\
&\quad + i2\pi \sum_{k=1}^K A_k(t) \phi'_k(t) e^{i2\pi\phi_k(t)} \frac{1}{\theta} h \left(\frac{\omega - \phi'_k(t)}{\theta} \right) \\
&\quad - \sum_{k=1}^K A_k(t) e^{i2\pi\phi(t)} \phi''_k(t) \frac{1}{\theta^2} h' \left(\frac{\omega - \phi'_k(t)}{\theta} \right) \\
&= \sum_{k=1}^K A'_k(t) e^{i2\pi\phi(t)} \frac{1}{\theta} h \left(\frac{\omega - \phi'_k(t)}{\theta} \right) \\
&\quad + i2\pi \sum_{k=1}^K A_k(t) \phi'_k(t) e^{i2\pi\phi_k(t)} \frac{1}{\theta} h \left(\frac{\omega - \phi'_k(t)}{\theta} \right) \\
&\quad + \partial_\omega \sum_{k=1}^K A_k(t) e^{i2\pi\phi(t)} \phi''_k(t) \frac{1}{\theta} h \left(\frac{\omega - \phi'_k(t)}{\theta} \right).
\end{aligned} \tag{15}$$

By the fact that $\omega \tilde{R}_f(t, \omega) = \sum_{k=1}^K A_k(t) \omega e^{i2\pi\phi_k(t)} \frac{1}{\theta} h \left(\frac{\omega - \phi'_k(t)}{\theta} \right)$, we have

$$\begin{aligned}
&\partial_t \tilde{R}_f(t, \omega) - i2\pi\omega \tilde{R}_f(t, \omega) \\
&= \sum_{k=1}^K A'_k(t) e^{i2\pi\phi(t)} \frac{1}{\theta} h \left(\frac{\omega - \phi'_k(t)}{\theta} \right) \\
&\quad - i2\pi \sum_{k=1}^K A_k(t) (\omega - \phi'_k(t)) e^{i2\pi\phi_k(t)} \frac{1}{\theta} h \left(\frac{\omega - \phi'_k(t)}{\theta} \right) \\
&\quad + \partial_\omega \sum_{k=1}^K A_k(t) e^{i2\pi\phi_k(t)} \phi''_k(t) \frac{1}{\theta} h \left(\frac{\omega - \phi'_k(t)}{\theta} \right).
\end{aligned} \tag{16}$$

We first discuss the case when $f \in \mathcal{A}_{\epsilon, d}^{c_1, c_2}$; that is, $|\phi''_k(t)| \leq \epsilon |\phi'_k(t)|$ for all $t \in \mathbb{R}$. Note that by the assumption of frequency separation (7) and the fact that $\theta \ll 1$, $[\phi'_l(t) - \theta\sigma, \phi'_l(t) + \theta\sigma] \cap [\phi'_k(t) - \theta\sigma, \phi'_k(t) + \theta\sigma] = \emptyset$ when $l \neq k$. Thus we have

$$\left| \sum_{k=1}^K A'_k(t) e^{i2\pi\phi_k(t)} \frac{1}{\theta} h \left(\frac{\omega - \phi'_k(t)}{\theta} \right) \right|^2 = \sum_{k=1}^K |A'_k(t)|^2 \frac{1}{\theta^2} h^2 \left(\frac{\omega - \phi'_k(t)}{\theta} \right). \tag{17}$$

Indeed, when $\omega \in [\phi'_l(t) - \theta\sigma, \phi'_l(t) + \theta\sigma]$, we have

$$\left| \sum_{k=1}^K A'_k(t) e^{i2\pi\phi(t)} \frac{1}{\theta} h \left(\frac{\omega - \phi'_k(t)}{\theta} \right) \right|^2 = |A'_l(t)|^2 \frac{1}{\theta^2} h^2 \left(\frac{\omega - \phi'_l(t)}{\theta} \right). \tag{18}$$

The same argument holds for the other terms on the right hand side of (16). As a result, by a direct calculation, for any non-empty finite interval $I \subset \mathbb{R}$, we have

$$\begin{aligned} & \left\| \sqrt{\theta} \left(\partial_t \tilde{R}_f(t, \omega) - i2\pi\omega \tilde{R}_f(t, \omega) \right) \right\|_{L^2(I \times [0, \infty))}^2 \\ & \leq \left(\epsilon^2 J_{0,0,2} + 2\pi\theta\epsilon J_{1,0,2} + 4\pi^2\theta^2 J_{2,0,2} + \frac{\epsilon^2 c_2^2}{\theta^2} J_{0,1,2} \right) c_2^2 I, \end{aligned} \quad (19)$$

where $J_{n,m,l} := \int \eta^n [\partial_\eta^m h(\eta)]^l d\eta$, where $n, m, l = 0, 1, \dots$. Thus, when ϵ is small enough, $\left\| \sqrt{\theta} \left(\partial_t \tilde{R}_f(t, \omega) - i2\pi\omega \tilde{R}_f(t, \omega) \right) \right\|_{L^2(I \times [0, \infty))}^2$ is small. Here, we mention that as the dynamic inside the signal we have interest is “momentary”, we would expect to have a small error between $\partial_t \tilde{R}_f(t, \omega)$ and $i2\pi\omega \tilde{R}_f(t, \omega)$ “locally”, which however, might accumulate when I becomes large. This observation leads to a variational approach discussed in [15]. Precisely, the authors in [15] considered to minimize the following functional when the signal f is observed on a non-empty finite interval I :

$$\begin{aligned} \mathcal{H}_0(F) &:= \int_I \left| \Re \int F(t, \omega) d\omega - f(t) \right|^2 dt \\ &+ \mu \iint_I |\partial_t F(t, \omega) - i2\pi\omega F(t, \omega)|^2 dt d\omega. \end{aligned} \quad (20)$$

The optimal F would be expected to approximate the iTFR of $f \in \mathcal{A}_{\epsilon,d}^{c_1, c_2, c_3}$ well. However, that optimization was not numerically carried out in [15].

Now we come back to the case we have interest; that is, $f \in \mathcal{Q}_{\epsilon,d}^{c_1, c_2, c_3}$. Since the condition on the CF terms, that is, $|\phi_k''(t)| \leq \epsilon |\phi_k'(t)|$, no longer holds, the above bound (19) does not hold and minimizing the functional \mathcal{H}_0 might not lead to the right solution. In this case, however, we still have the following bound by the same argument as that of (19):

$$\begin{aligned} & \left\| \sqrt{\theta} \left(\partial_t \tilde{R}_f(t, \omega) - i2\pi\omega \tilde{R}_f(t, \omega) - \partial_\omega \sum_{k=1}^K A_k(t) e^{i2\pi\phi_k(t)} \phi_k''(t) \frac{1}{\theta} h \left(\frac{\omega - \phi_k'(t)}{\theta} \right) \right) \right\|_{L^2(I \times [0, \infty))}^2 \\ & \leq \left\| \sqrt{\theta} \sum_{k=1}^K (A_k'(t) - i2\pi A_k(t)(\omega - \phi_k'(t))) e^{i2\pi\phi_k(t)} \frac{1}{\theta} h \left(\frac{\omega - \phi_k'(t)}{\theta} \right) \right\|_{L^2(I \times [0, \infty))}^2 \\ & \leq (\epsilon^2 J_{0,0,2} + 2\pi\theta\epsilon J_{1,0,2} + 4\pi^2\theta^2 J_{2,0,2}) c_2^2 I, \end{aligned} \quad (21)$$

Thus, once we find a way to express the extra term $\partial_\omega \sum_{k=1}^K A_k(t) e^{i2\pi\phi_k(t)} \phi_k''(t) \frac{1}{\theta} h \left(\frac{\omega - \phi_k'(t)}{\theta} \right)$ in a convenient formula, we could introduce another conditions on F .

In the special case when $K = 1$; that is, $f = A(t) \cos(2\pi\phi(t))$, we know that

$$\partial_\omega \left[A(t) e^{i2\pi\phi(t)} \phi''(t) \frac{1}{\theta} h \left(\frac{\omega - \phi'(t)}{\theta} \right) \right] = \phi''(t) \partial_\omega \tilde{R}_f(t, \omega). \quad (22)$$

Thus, we have

$$\theta \iint_I |\partial_t \tilde{R}_f(t, \omega) - i2\pi\omega \tilde{R}_f(t, \omega) + \phi''(t) \partial_\omega \tilde{R}_f(t, \omega)|^2 dt d\omega = O(\theta^2, \theta\epsilon, \epsilon^2). \quad (23)$$

Thus, we could consider the following functional

$$\theta \iint |\partial_t F(t, \omega) - i2\pi\omega F(t, \omega) + \alpha(t)\partial_\omega F(t, \omega)|^2 dt d\omega, \quad (24)$$

where $\alpha(t) \in \mathbb{R}$ is used to capture the CF term associated with the “fast varying instantaneous frequency”. Thus, when $K = 1$, we can capture more general oscillatory signals by considering the following functional when the signal is observed on a non-empty finite interval $I \subset \mathbb{R}$:

$$\begin{aligned} \mathcal{H}(F, \alpha) := & \int_I \left| \Re \int F(t, \omega) d\omega - f(t) \right|^2 dt \\ & + \mu\theta \iint_I |\partial_t F(t, \omega) - i2\pi\omega F(t, \omega) + \alpha(t)\partial_\omega F(t, \omega)|^2 dt d\omega \\ & + \lambda \|F\|_{L^1(I \times \mathbb{R})} + \gamma \|\alpha\|_{L^2(I \times \mathbb{R})}, \end{aligned} \quad (25)$$

where $F \in L^2(I \times \mathbb{R})$ is the function defined on the TF plane restricted on $I \times \mathbb{R}$. Note that the L^1 norm is another constraint we introduce in order to enhance the sharpness of the TF representation. Indeed, we would expect to introduce a sparse TF representation when the signal is composed of several gIMT.

In general when $K > 1$, we cannot link $\partial_\omega \sum_{k=1}^K A_k(t) e^{i2\pi\phi_k(t)} \phi_k''(t) \frac{1}{\theta} h\left(\frac{\omega - \phi_k'(t)}{\theta}\right)$ to $\partial_\omega \tilde{R}_f(t, \omega)$ by any function on t like that in (22). In this case, we could expect to find another function $G \in L^2(I \times \mathbb{R})$ so that

$$G(t, \omega) = \begin{cases} \phi_k''(t) & \text{when } \omega \in [\phi_k'(t) - \theta\sigma, \phi_k'(t) + \theta\sigma] \\ 0 & \text{otherwise.} \end{cases} \quad (26)$$

and hence $G(t, \omega) \partial_\omega \tilde{R}_f(t, \omega) = \partial_\omega \sum_{k=1}^K A_k(t) e^{i2\pi\phi_k(t)} \phi_k''(t) \frac{1}{\theta} h\left(\frac{\omega - \phi_k'(t)}{\theta}\right)$. Thus, we could consider minimizing the following functional for a given function f observed on a non-empty finite interval $I \subset \mathbb{R}$:

$$\begin{aligned} \mathcal{H}(F, G) := & \int_I \left| \Re \int F(t, \omega) d\omega - f(t) \right|^2 dt \\ & + \mu\theta \iint_I |\partial_t F(t, \omega) - i2\pi\omega F(t, \omega) + G(t, \omega)\partial_\omega F(t, \omega)|^2 dt d\omega \\ & + \lambda \|F\|_{L^1(I \times \mathbb{R})} + \frac{\gamma}{\sqrt{\theta}} \|G\|_{L^2(I \times \mathbb{R})}. \end{aligned} \quad (27)$$

Here, the L^2 penalty term $\|G\|_{L^2}$ has $1/\sqrt{\theta}$ in front of it since

$$\|G\|_{L^2(I \times [0, \infty))} = \sqrt{2\theta\sigma} \sum_{k=1}^K \|\phi_k''\|_{L^2(I \times [0, \infty))}. \quad (28)$$

Thus, the L^2 penalty term does not depend on θ . It is also clear that the L^1 penalty term in the above functional does not depend on θ as we have

$$\int \int_I \left| \sum_{k=1}^K A_k(t) e^{i2\pi\phi_k(t)} \frac{1}{\theta} h\left(\frac{\omega - \phi_k'(t)}{\theta}\right) \right| dt d\omega = \sum_{k=1}^K \|A_k(t)\|_{L^1(I)}. \quad (29)$$

4. Numerical Algorithm

We consider the following functionals associated with (25):

$$\begin{aligned}\mathcal{H}(F, \alpha) &= \int_{\mathbb{R}} \left| \Re \int_{\mathbb{R}} F(t, \omega) d\omega - f(t) \right|^2 dt \\ &\quad + \tilde{\mu} \left(\tilde{\lambda} \iint_{\mathbb{R}} |\partial_t F(t, \omega) - i2\pi\omega F(t, \omega) + \alpha(t) \partial_\omega F(t, \omega)|^2 dt d\omega + (1 - \tilde{\lambda}) \|F\|_{L^1} \right) + \gamma \|\alpha\|_{L^2}^2 \\ &= \mathcal{G}(F, \alpha) + \Psi(F, \alpha),\end{aligned}\tag{30}$$

where

$$\begin{aligned}\mathcal{G}(F, \alpha) &:= \int_{\mathbb{R}} \left| \Re \int_{\mathbb{R}} F(t, \omega) d\omega - f(t) \right|^2 dt \\ &\quad + \tilde{\mu} \tilde{\lambda} \iint_{\mathbb{R}} |\partial_t F(t, \omega) - i2\pi\omega F(t, \omega) + \alpha(t) \partial_\omega F(t, \omega)|^2 dt d\omega,\end{aligned}\tag{31}$$

$$\Psi(F, \alpha) := \tilde{\mu}(1 - \tilde{\lambda}) \|F\|_{L^1} + \gamma \|\alpha\|_{L^2}^2,\tag{32}$$

t is the time and ω is the frequency. The numerical implementation of (27) follows the same lines while we have to discretize a two dimensional function G . Compared to (25), we have redefined the role of the hyperparameter μ and λ . Here, $\tilde{\mu} \in \mathbb{R}_+$ balance between the data fidelity term $\int_{\mathbb{R}} \left| \Re \int_{\mathbb{R}} F(t, \omega) d\omega - f(t) \right|^2 dt$ allowing the reconstruction, and the regularization term which controls the variation on the derivatives, and the sparsity of the solution. The parameter $\tilde{\lambda} \in [0, 1]$ allows one to balance between the sparsity prior and the constraint of the derivatives. This choice will simplify the choice of the regularization parameters. Clearly, by setting $\mu = \tilde{\mu} \tilde{\lambda}$ and $\lambda = \tilde{\mu}(1 - \tilde{\lambda})$ we recover the original formulation (25).

4.1. Numerical discretization

Numerically, we consider the following discretization of F by taking $\Delta_t > 0$ and $\Delta_\omega > 0$ as the sampling periods in the time axis and frequency axis. We also restrict F to time $[0, M\Delta_t]$ and to the frequencies $[-N\Delta_\omega, N\Delta_\omega]$. Then, we discretize F as $\mathbf{F} \in \mathbb{C}^{(N+1) \times (M+1)}$ and α as $\boldsymbol{\alpha} \in \mathbb{R}^{M+1}$, where

$$\mathbf{F}_{n,m} = F(t_m, \omega_n), \quad \boldsymbol{\alpha}_m = \alpha(t_m),\tag{33}$$

$t_m := m\Delta_t$, $\omega_n := n\Delta_\omega$, $n = -N, \dots, N$ and $m = 0, 1, \dots, M$. The observed signal $f(t)$ is discretized as a $(M+1)$ -dim vector \mathbf{f} , where

$$\mathbf{f}_l = f(t_l).\tag{34}$$

Note that the sampling period of the signal Δ_t and M most of time are determined by the data collection procedure. We could set $\Delta_\omega = \frac{1}{M\Delta_t}$ and $N = \lceil M/2 \rceil$ suggested by the Nyquist rate in the sampling theory.

Next, using the rectangle method, we could discretize $\mathcal{G}(F, \alpha)$ directly by

$$\begin{aligned} \mathcal{G}(\mathbf{F}, \boldsymbol{\alpha}) := & \sum_{m=0}^M \left| \sum_{n=-N}^N 2\Re(F(t_m, \omega_n)) \Delta_\omega - f(t_m) \right|^2 \Delta_t \\ & + \mu \sum_{m=0}^M \sum_{n=-N}^N |\partial_t F(t_m, \omega_n) - i2\pi\omega_n F(t_m, \omega_n) + \alpha(t_m) \partial_\omega F(t_m, \omega_n)|^2 \Delta_t \Delta_\omega. \end{aligned} \quad (35)$$

The partial derivative $\partial_t F$ can be implemented by the straight finite difference; that is, take a $(M+1) \times (M+1)$ finite difference matrix \mathbf{D}_{M+1} so that $\mathbf{F}\mathbf{D}_{M+1}$ approximates the discretization of $\partial_t F$. However, this choice may lead to numerical instability. Instead, one can implement the partial derivative in the Fourier domain, using that $\partial_t F(t_m, \omega_n) = \mathcal{F}^{-1} \left(i2\pi\xi_k \hat{F}(\xi_k, \omega_n) \right) [m]$, where $\hat{F} = \mathcal{F}(F)$ and \mathcal{F} denotes the finite Fourier transform. For the sake of simplicity, we still denote by ∂_t or ∂_ω the discretization operator in the discret domain, whatever the chosen method (finite difference or in the Fourier domain). Also denote $\mathbf{1} = (1, \dots, 1)^T \in \mathbb{R}^{M+1}$. In the matrix form, the functional $\mathcal{G}(F, \alpha)$ is thus discretized as

$$\mathcal{G}(\mathbf{F}, \boldsymbol{\alpha}) = \Delta_t \|\mathcal{A}\mathbf{F} - \mathbf{f}\|^2 + \Delta_t \Delta_\omega \mu \|\mathcal{B}(\mathbf{F}, \boldsymbol{\alpha})\|^2, \quad (36)$$

where

$$\begin{aligned} \mathcal{A}: \quad \mathbb{C}^{(M+1) \times (M+1)} & \rightarrow \mathbb{R}^{M+1} \\ \mathbf{F} & \mapsto 2\Re(\mathbf{1}^T \mathbf{F}) \Delta_\omega, \end{aligned} \quad (37)$$

$$\begin{aligned} \mathcal{B}: \quad \mathbb{C}^{(M+1) \times (M+1)} \times \mathbb{C}^{M+1} & \rightarrow \mathbb{C}^{(N+1) \times (M+1)} \\ (\mathbf{F}, \boldsymbol{\alpha}) & \mapsto \partial_t \mathbf{F} - i2\pi\boldsymbol{\omega} \mathbf{F} + \partial_\omega \mathbf{F} \text{diag}(\boldsymbol{\alpha}), \end{aligned} \quad (38)$$

and $\boldsymbol{\omega} = \text{diag}(-N\Delta_\omega, \dots, 0, \Delta_\omega, 2\Delta_\omega, \dots, N\Delta_\omega) \in \mathbb{R}^{(M+1) \times (M+1)}$.

4.2. Expression of the gradient operator

Denote $\mathcal{G}_\alpha(\mathbf{F}) := \mathbf{F} \mapsto \mathcal{G}(\mathbf{F}, \boldsymbol{\alpha})$ and $\mathcal{B}_\alpha(\mathbf{F}) := \mathbf{F} \mapsto \mathcal{B}(\mathbf{F}, \boldsymbol{\alpha})$; that is, $\boldsymbol{\alpha}$ is fixed. Similarly, define $\mathcal{G}_\mathbf{F}(\boldsymbol{\alpha}) := \boldsymbol{\alpha} \mapsto \mathcal{G}(\mathbf{F}, \boldsymbol{\alpha})$ and $\mathcal{B}_\mathbf{F}(\boldsymbol{\alpha}) := \boldsymbol{\alpha} \mapsto \mathcal{B}(\mathbf{F}, \boldsymbol{\alpha})$; that is, \mathbf{F} is fixed. We will evaluate the gradient of \mathcal{G}_α and $\mathcal{G}_\mathbf{F}$ after discretization for the gradient decent algorithm. Take $\mathbf{G} \in \mathbb{C}^{(M+1) \times (M+1)}$. The gradient of \mathcal{G}_α after discretization is evaluated by

$$\begin{aligned} \nabla \mathcal{G}_\alpha|_{\mathbf{F}} \mathbf{G} &= \lim_{h \rightarrow 0} \frac{\mathcal{G}_\alpha(\mathbf{F} + h\mathbf{G}) - \mathcal{G}_\alpha(\mathbf{F})}{h} \\ &= 2\Delta_t (\mathcal{A}\mathbf{F} - \mathbf{f})^T \mathcal{A}\mathbf{G} + 2\Delta_t \Delta_\omega \mu \langle \mathcal{B}_\alpha \mathbf{F}, \mathcal{B}_\alpha \mathbf{G} \rangle \\ &= \langle 2\Delta_t \mathcal{A}^* (\mathcal{A}\mathbf{F} - \mathbf{f}) + 2\Delta_t \Delta_\omega \mu \mathcal{B}_\alpha^* \mathcal{B}_\alpha \mathbf{F}, \mathbf{G} \rangle. \end{aligned} \quad (39)$$

As a result, we have

$$\nabla \mathcal{G}_\alpha|_{\mathbf{F}} = 2\Delta_t \mathcal{A}^* (\mathcal{A}\mathbf{F} - \mathbf{f}) + 2\Delta_t \Delta_\omega \mu \mathcal{B}_\alpha^* \mathcal{B}_\alpha \mathbf{F}. \quad (40)$$

where \mathcal{A}^* and \mathcal{B}_α^* are adjoint operators of \mathcal{A} and \mathcal{B}_α respectively. Now we expand \mathcal{A}^* and \mathcal{B}_α^* . Take $\mathbf{g} \in \mathbb{R}^{M+1}$. We have

$$\begin{aligned}\langle \mathcal{A}\mathbf{F}, \mathbf{g} \rangle &= \sum_{m=0}^M \left(\sum_{n=-N}^N 2\Re \mathbf{F}_{n,m} \Delta_\omega \right) \mathbf{g}_m \\ &= \sum_{m=0}^M \sum_{n=-N}^N 2\Re \mathbf{F}_{n,m} \Re(\Delta_\omega \mathbf{g}_m),\end{aligned}\tag{41}$$

and

$$\begin{aligned}\langle \mathbf{F}, \mathcal{A}^* \mathbf{g} \rangle &= \sum_{m=0}^M \sum_{n=-N}^N \mathbf{F}_{n,m} \overline{(\mathcal{A}^* \mathbf{g})_{n,m}} \\ &= \sum_{m=0}^M \sum_{n=-N}^N 2\Re \mathbf{F}_{n,m} \Re(\mathcal{A}^* \mathbf{g})_{n,m} + \sum_{m=0}^M \sum_{n=-N}^N \Im \mathbf{F}_{n,m} \Im(\mathcal{A}^* \mathbf{g})_{n,m} \\ &\quad + i \sum_{m=0}^M \sum_{n=-N}^N \Im \mathbf{F}_{n,m} 2\Re(\mathcal{A}^* \mathbf{g})_{n,m} - i \sum_{m=0}^M \sum_{n=-N}^N \Re \mathbf{F}_{n,m} \Im(\mathcal{A}^* \mathbf{g})_{n,m}.\end{aligned}\tag{42}$$

Since $\langle \mathcal{A}\mathbf{F}, \mathbf{g} \rangle = \langle \mathbf{F}, \mathcal{A}^* \mathbf{g} \rangle$ for all \mathbf{F} and \mathbf{g} , we conclude that

$$\begin{aligned}\mathcal{A}^* : \mathbb{R}^{M+1} &\rightarrow \mathbb{C}^{(M+1) \times (M+1)} \\ \mathbf{g} &\mapsto 2\Delta_\omega \begin{pmatrix} \mathbf{g}_1 & \cdots & \mathbf{g}_{M+1} \\ \vdots & & \vdots \\ \mathbf{g}_1 & \cdots & \mathbf{g}_{M+1} \end{pmatrix}.\end{aligned}\tag{43}$$

To calculate \mathcal{B}_α^* , by a direct calculation we have

$$\begin{aligned}\langle \mathcal{B}_\alpha \mathbf{F}, \mathbf{G} \rangle &= \langle \partial_t \mathbf{F} - i2\pi\omega \mathbf{F} + \partial_\omega \mathbf{F} \text{diag}(\boldsymbol{\alpha}), \mathbf{G} \rangle \\ &= \langle \mathbf{F}, -\partial_t \mathbf{G} + i2\pi\omega \mathbf{G} - \partial_\omega \mathbf{G} \text{diag}(\boldsymbol{\alpha}) \rangle \\ &= \langle \mathbf{F}, \mathcal{B}_\alpha^* \mathbf{G} \rangle,\end{aligned}\tag{44}$$

where $\mathbf{G} \in \mathbb{C}^{(M+1) \times (M+1)}$. Thus, we conclude that

$$\begin{aligned}\mathcal{B}_\alpha^* : \mathbb{C}^{(M+1) \times (M+1)} &\rightarrow \mathbb{C}^{(M+1) \times (M+1)} \\ \mathbf{G} &\mapsto -\partial_t \mathbf{G} + i2\pi\omega \mathbf{G} - \partial_\omega \mathbf{G} \text{diag}(\boldsymbol{\alpha}).\end{aligned}\tag{45}$$

As a result, the first part of $\nabla \mathcal{G}_\alpha|_{\mathbf{F}}, 2\Delta_t \mathcal{A}^*(\mathcal{A}\mathbf{F} - \mathbf{f})$, can be numerically expressed as

$$4\Delta_t \Delta_\omega \begin{pmatrix} \Delta_\omega \Re \sum_{n=0}^N \mathbf{F}_{n,1} - \mathbf{f}_1 & \cdots & \Delta_\omega \Re \sum_{n=0}^N \mathbf{F}_{n,M+1} - \mathbf{f}_{M+1} \\ \vdots & & \vdots \\ \Delta_\omega \Re \sum_{n=0}^N \mathbf{F}_{n,1} - \mathbf{f}_1 & \cdots & \Delta_\omega \Re \sum_{n=0}^N \mathbf{F}_{n,M+1} - \mathbf{f}_{M+1} \end{pmatrix} \in \mathbb{R}^{(M+1) \times (M+1)}.\tag{46}$$

and the second term

$$2\Delta_t\Delta_\omega\mu\mathcal{B}^*\mathcal{B}\mathbf{F} = 2\Delta_t\Delta_\omega\mu\left(-\partial_t\partial_t\mathbf{F} + i4\pi\omega\partial_t\mathbf{F} - \partial_t\partial_\omega\mathbf{F}\text{diag}(\boldsymbol{\alpha}) + 4\pi^2\omega^2\mathbf{F} \right. \\ \left. + i2\pi\omega\partial_\omega\mathbf{F}\text{diag}(\boldsymbol{\alpha}) - \partial_\omega\partial_t\mathbf{F}\text{diag}(\boldsymbol{\alpha}) + i2\pi\partial_\omega\omega\mathbf{F}\text{diag}(\boldsymbol{\alpha}) - \partial_\omega\partial_\omega\mathbf{F}\text{diag}(\boldsymbol{\alpha})\right). \quad (47)$$

Similarly, by taking $\beta \in \mathbb{C}^{M+1}$, the gradient of $\mathcal{G}_{\mathbf{F}}$ at $\boldsymbol{\alpha}$ after discretization is evaluated by

$$\begin{aligned} \nabla\mathcal{G}_{\mathbf{F}}|_{\boldsymbol{\alpha}}\beta &= \lim_{h \rightarrow 0} \frac{\mathcal{G}_{\mathbf{F}}(\boldsymbol{\alpha} + h\beta) - \mathcal{G}_{\mathbf{F}}(\boldsymbol{\alpha})}{h} \\ &= \text{tr}((\partial_\omega\mathbf{F}\text{diag}(\beta))^*(\partial_\omega\mathbf{F}\text{diag}(\boldsymbol{\alpha}))) \\ &= -\beta^*\text{tr}(\mathbf{F}^*\partial_\omega\partial_\omega\mathbf{F}\text{diag}(\boldsymbol{\alpha})). \end{aligned} \quad (48)$$

Thus, we have

$$\nabla\mathcal{G}_{\mathbf{F}}|_{\boldsymbol{\alpha}} = -\text{tr}(\mathbf{F}^*\partial_\omega\partial_\omega\mathbf{F}\text{diag}(\boldsymbol{\alpha})) \in \mathbb{C}^{M+1}, \quad (49)$$

where

$$(\nabla\mathcal{G}_{\mathbf{F}}|_{\boldsymbol{\alpha}})_m = \alpha(t_m) \sum_{n=-N}^N [\partial_\omega F(t_m, \omega_n)]^2. \quad (50)$$

4.3. Minimize the functional $\mathcal{H}(F, \alpha)$

We now have all the results needed to propose an optimization algorithm to minimize the functional $\mathcal{H}(F, \alpha)$. The minimization of $\mathcal{H}(F, G)$ in (27) is the same so we skip it. The functional we would like to minimize depends on two terms, F and α . While the PALM algorithm studied in [6] provides a simple procedure to minimize (25), this algorithm appeared to be too slow in practice for this problem. Since the functional spaces F and α live are convex, we will therefore minimize the functional alternately by optimizing one of these two terms when the other one is fixed; that is,

$$\begin{cases} F_{k+1} = \arg \min_F \mathcal{H}(F, \alpha_k) \\ \alpha_{k+1} = \arg \min_\alpha \mathcal{H}(F_{k+1}, \alpha). \end{cases} \quad (51)$$

with $\alpha_0 = 0$ and $F_0 = 0$ are used to initialize the algorithm. A discussion on convergence results of this classical Gauss-Seidel method can be found in [6].

As we will see in next subsections, if we can reach the global minimizer of $\alpha \mapsto \mathcal{H}(F_{k+1}, \alpha)$, finding a minimizer of $F \mapsto \mathcal{H}(F, \alpha_k)$ requires the use of an iterative algorithm. We provide in Appendix C a convergence result of the practical algorithm we propose.

4.4. Minimization of $\mathcal{H}_\alpha := \mathcal{H}(\cdot, \alpha)$

When α is fixed, \mathcal{H}_α is a convex non smooth functional, involving a convex and Lipschitz differentiable term (the function $\mathcal{G}_\alpha := \mathcal{G}(\cdot, \alpha)$), and a convex but non-smooth term (the $\Psi_\alpha := \Psi(\cdot, \alpha)$ regularizer). Popular proximal algorithms such as forward-backward [14] or the Fast Iterative Shrinkage/Thresholding Algorithm (FISTA) [5, 7] can

then be employed. FISTA has the great advantage to reach the optimal rate of convergence; that is, if $\check{\mathbf{F}}$ is the convergence point, $\mathcal{H}_\alpha(\mathbf{F}_k) - \mathcal{H}_\alpha(\check{\mathbf{F}}) = \mathcal{O}\left(\frac{1}{k^2}\right)$, while the forward-backward procedure converge in $\mathcal{O}\left(\frac{1}{k}\right)$ (see [48] for a great review of proximal methods and their acceleration). This speed of convergence is usually observed in practice [38], and has been confirmed in our experiments (not shown in this paper). Contrary to the forward-backward, one limitation of the original FISTA [5] is that the convergence is proven only on the sequence $(\mathcal{H}_\alpha(\mathbf{F}_k))_k$ rather than on the iterates $(\mathbf{F}_k)_k$. However, the latest study [7] gives a version of FISTA, which fills in this gap while maintaining the same convergence rate. As far as we know, it is the only algorithm with these two properties, and then will be use in the following. Yet another shortcoming of the original FISTA is that the algorithm does not produce a monotonic decreasing of the functional, but a monotonic version is available [4] and is used in this paper.

In short, FISTA relies on three steps

1. A gradient descent step on the smooth term \mathcal{G}_α ;
2. A soft-shrinkage operation, known as the proximal step;
3. A relaxation step.

The algorithm is summarized in Algorithm 1. In practice, the Lipschitz constant can be evaluated using a classical power iteration procedure, or using a backtracking step inside the algorithm (see [5] for details). $\nabla \mathcal{G}_\alpha$ is given by Eq. (46) and Eq. (47).

Moreover, when the signal \mathbf{f} is real and α is real, we can limit the optimization to the positive frequencies such that $\mathbf{F} \in \mathbb{C}^{(N+1) \times (M+1)}$, with $N = \lceil M/2 \rceil$. Indeed, one can show that there exists a solution \mathbf{F} which has an Hermitian symmetry properties, i.e. such that $F(t, \omega) = \overline{F(t, -\omega)}$. In order to prove this result, we remark that we have

$$\overline{\nabla \mathcal{G}_\alpha|_{\mathbf{F}}(t, -\omega)} = \nabla \mathcal{G}_\alpha|_{\overline{\mathbf{F}}}(t, -\omega), \quad (52)$$

which can be easily checked thanks to Eq. (46) and (47). Then, if \mathbf{F}_0 is Hermitian symmetric, one can prove by induction that at each iteration, \mathbf{F}_k is Hermitian symmetric.

4.5. Minimization of $\mathcal{H}_F := \mathcal{H}(F, \cdot)$

Once \mathbf{F}_k is estimated, the minimization of $\mathcal{H}_{\mathbf{F}_k}$ reduces to a simple quadratic minimization:

$$\alpha_{k+1} = \underset{\alpha}{\operatorname{argmin}} \left\{ \mu \sum_{m=0}^M \sum_{n=0}^N |\partial_t F(t_m, \omega_n) - i2\pi\omega_n F(t_m, \omega_n) + \alpha(t_m) \partial_\omega F(t_m, \omega_n)|^2 + \gamma \sum_{m=0}^M |\alpha(t_m)|^2 \right\}. \quad (53)$$

Thus, α can be estimated in a closed form as, for all $m = 0, \dots, M$,

$$\alpha_{k+1}(t_m) = \frac{2 \sum_{n=0}^N \Re \left(\partial_\omega \overline{F(t_m, \omega_n)} [\partial_t F(t_m, \omega_n) - i2\pi\omega_n F(t_m, \omega_n)] \right)}{\sum_{n=0}^N |\partial_\omega F(t_m, \omega_n)|^2 + \gamma/\mu}. \quad (54)$$

Algorithm 1 FISTA algorithm for \mathcal{H}_α : $\mathbf{F} = \text{FISTA}(\mathbf{F}_0, \alpha, \epsilon)$

Choose a stopping value ϵ .

The initial values are $\mathbf{F}_0 \in \mathbb{C}^{(N+1) \times (M+1)}$, $z_0 = \mathbf{F}_0$

Evaluate the Lipschitz constant $L = \|\nabla \mathcal{G}_\alpha\|^2$ by power iterations.

while $\frac{\|\mathbf{F}_{k+1} - \mathbf{F}_k\|}{\|\mathbf{F}_k\|} > \epsilon$ **do**

 Gradient step: $\mathbf{F}_{k+1/2} \leftarrow z_k - \frac{1}{L} \nabla \mathcal{G}_\alpha|_{z_k}$ (see (46) and (47));

 Proximal step: $\mathbf{F}_{k+1/2} \leftarrow \mathbf{F}_{k+1/2} \left(1 - \frac{\lambda/L}{\|\mathbf{F}_{k+1/2}\|}\right)^+$;

 Monotonic step:

if $\mathcal{H}(\mathbf{F}_{k+1/2}, \alpha) < \mathcal{H}(\mathbf{F}_k, \alpha)$ **then**

$\mathbf{F}_{k+1} = \mathbf{F}_{k+1/2}$

else

$\mathbf{F}_{k+1} = \mathbf{F}_k$

end if

 Relaxation step: $z_{k+1} \leftarrow \mathbf{F}_{k+1} + \frac{k}{k+2}(\mathbf{F}_{k+1} - \mathbf{F}_k) + \frac{k+1}{k+2}(\mathbf{F}_{k+1/2} - \mathbf{F}_k)$;

$k = k + 1$;

end while

Output \mathbf{F} .

Algorithm 2 Algorithm for minimization of \mathcal{H}

Choose a stopping value ϵ_1 for the FISTA algorithm;

Choose a stopping value ϵ_2 for the alternating minimization;

Choose a set of decreasing values $I_{\tilde{\mu}}$ for the parameter $\tilde{\mu} \in \mathbb{R}_+$.

Choose the parameters $\tilde{\lambda} \in [0, 1]$ and $\gamma \in \mathbb{R}_+$;

The initial values are $k = 0$, $\mathbf{F}_0 = \mathbf{0}$, $\alpha_0 = \mathbf{0}$;

for $\tilde{\mu} \in I_{\tilde{\mu}}$ **do**

while $\frac{\|\mathbf{F}_{k+1} - \mathbf{F}_k\|}{\|\mathbf{F}_k\|} > \epsilon_1$ **do**

 FISTA step: $\mathbf{F}_{k+1} = \text{FISTA}(\mathbf{F}_k, \alpha_k, \epsilon_1)$ (see Alg. 1);

 alpha estimation step (see Eq. (54));

$k = k + 1$;

end while

end for

Output \mathbf{F}, α ;

4.6. General algorithm

We summarize in Algorithm 2 the practical procedure to minimize \mathcal{H} (25). The choices of the parameters are discussed below.

- **Stopping criterion.** As the functional $F \mapsto \mathcal{H}(\mathcal{F}, \alpha)$ is convex, a good stopping criterion for FISTA is the so-called *duality gap*. However, the duality gap cannot be computed easily here. We then choose the classical quantity, $\frac{\|\mathbf{F}_{k+1} - \mathbf{F}_k\|}{\|\mathbf{F}_k\|}$, to stop the FISTA inner loop as well as the alternating algorithm; that is, the algorithm stops when both the stopping criteria, $\frac{\|\mathbf{F}_{k+1} - \mathbf{F}_k\|}{\|\mathbf{F}_k\|} \leq \epsilon_1$ and $\frac{\|\alpha_{k+1} - \alpha_k\|}{\|\alpha_k\|} \leq \epsilon_2$ for the chosen $\epsilon_1, \epsilon_2 \geq 0$, are satisfied. ϵ_1 and ϵ_2 can be set to 5×10^{-4} in practice: smaller

value produce a much slower algorithm for similar results.

- **Set of values $I_{\tilde{\mu}}$.** A practical choice of $I_{\tilde{\mu}}$ is a set of K values uniformly distributed on the logarithmic scale. In the noise free case, one must choose a sufficiently small $\tilde{\mu}$. However, a small value of $\tilde{\mu}$ gives a very slow algorithm. A practical strategy is to use a fixed point continuation [25], also known as warm start, strategy to minimize \mathcal{H} . If the noise is taken into account, the final $\tilde{\mu}$ cannot be known in advance, but can be chosen to be the one leading to the best result among the K obtained minimizers. Here, we choose $\tilde{\mu}$ according to the discrepancy principle [40]. Another approach could be the GSURE approach [19] (not derived in this work).
- **Parameter $\tilde{\lambda}$.** This parameter must be chosen between 0 and 1. The closer $\tilde{\lambda}$ is to 1, the more importance is given to the constraints on the derivatives. As these constraints should be satisfied as much as possible, we choose in practice $\tilde{\lambda} \simeq 0.99$.
- **Parameter γ .** The influence of this parameter is not dominant on the results. We set $\gamma \simeq 10^{-3}$ in order to prevent any division by 0 during the estimation of α by (54).
- **Initialization of the algorithm.** The choice of $\alpha = \mathbf{0}$ appears to be natural, as we cannot have access to the chirp factor. The first iteration of Algorithm 2 is equivalent to an estimation without taking this chirp factor into account. However, this initialization can have some influence on the speed of the algorithm [5]. As the solution is expected to be sparse, $\mathbf{F} = \mathbf{0}$ seems to be a reasonable choice.

5. Numerical Results

In this section we show numerical simulation results of the proposed algorithm. The code and simulated data are available via request. In this section, we take W to be the standard Brownian motion defined on $[0, \infty)$ and define a *smoothed Brownian motion with bandwidth $\sigma > 0$* as

$$\Phi_{\sigma} := W \star K_{\sigma}, \quad (55)$$

where K_{σ} is the Gaussian function with the standard deviation $\sigma > 0$ and \star denotes the convolution operator.

5.1. Single component, noise-free

The first example is a semi-real example which is inspired from a medical challenge. Atrial fibrillation (Af) is a pathological condition associated with high mortality and morbidity [32]. It is well known that the subject with Af would have irregularly irregular heart beats. In the language under our framework, the instantaneous frequency of the electrocardiogram signal recorded from an Af patient varies fast. To study this kind of signal with fast varying instantaneous frequency, we pick a patient with Af and determine its instantaneous heart rate by evaluating its R peak to R peak intervals. Precisely, if the R peaks are located on t_i , we generate a non-uniform sampling of the instantaneous heart rate and denote it as $(t_i, 1/(t_{i+1} - t_i))$. Then the instantaneous heart rate, denoted as

$\phi_1'(t)$, is approximated by the cubic spline interpolation. Next, define another a random process A_1 on $[0, L]$ by

$$A_1(t) = 1 + \frac{\Phi_{\sigma_1}(t) + \|\Phi_{\sigma_1}\|_{L^\infty[0,L]}}{2\|\Phi_{\sigma_1}\|_{L^\infty[0,L]}}, \quad (56)$$

where $t \in [0, L]$ and $\sigma_1 > 0$. Note that A_1 is a positive random process and in general there is no close form expression of $A_1(t)$ and $\phi_1(t)$. The dynamic of both components can be visually seen from the signal. We then generate an oscillatory signal with fast varying instantaneous frequency

$$f_1(t) = A_1(t) \cos(2\pi\phi_1(t)), \quad (57)$$

where $A_1(t)$ is a realization of the random process defined in (56). We take $L = 80$, sample f_1 with the sampling rate $\Delta t = 1/10$, $\sigma_1 = 100$, $\sigma_2 = 200$. To compare the result with other methods, in addition to showing the result of the proposed algorithm, we also show the analysis results of STFT and synchrosqueezed STFT. In the STFT and synchrosqueezed STFT, we take the window function g as a Gaussian function with the standard deviation $\sigma = 1$. See Figure 1 for the result. We mention that in this section, when we plot the tvPS, we compress its dynamical range by the following procedure. Denoted the discretized tvPS as $\mathfrak{R} \in \mathbb{R}^{m \times n}$, where $m, n \in \mathbb{N}$ stand for the number of discrete frequencies and the number of time samples, respectively. Set M to be the 99.9% quantile of the absolute values of all entries of \mathfrak{R} , then normalize the discretized tvPS by M , and obtain $\tilde{\mathfrak{R}} \in \mathbb{R}^{m \times n}$ so that $\tilde{\mathfrak{R}}(i, j) := \max\{M, \mathfrak{R}(i, j)\}$ for $i = 1, \dots, m$ and $j = 1, \dots, n$. Then plot a gray-scale visualization of $\tilde{\mathfrak{R}}$ in the linear scale. From the figure, we see that the proposed algorithm Tycoon could extract this kind of fast varying IF well visually. However, although there are some periods where STFT and synchrosqueezed STFT show a dominant curve following the IF well, in general the IF information is blurred in their TF representations. In addition, by Tycoon, the chirp factor can be approximated up to some extent.

5.2. Two components, noise-free

In the second example, we consider an oscillatory signal with two gIMTs. Define random processes $A_2(t)$ and $\phi_2(t)$ on $[0, L]$ by

$$\begin{aligned} A_2(t) &= 1 + \frac{\Phi_{\sigma_1}(t) + 2\|\Phi_{\sigma_1}\|_{L^\infty[0,L]}}{3\|\Phi_{\sigma_1}\|_{L^\infty[0,L]}}, \\ \phi_2(t) &= \pi t + \int_0^t \left[\frac{\Phi_{\sigma_2}(s) + 0.5\|\Phi_{\sigma_2}\|_{L^\infty[0,L]}}{1.5\|\Phi_{\sigma_2}\|_{L^\infty[0,L]}} - \sin(s) \right] ds, \end{aligned} \quad (58)$$

where $t \in [0, L]$ and $\sigma_2 > 0$. Note that by definition ϕ_2 are both monotonically increasing random processes. The signal is constructed as

$$f(t) = f_1(t) + f_2(t), \quad (59)$$

where $f_2(t) = A_2(t) \cos(2\pi\phi_2(t))\chi_{[20,80]}(t)$ and χ is the indicator function. Again, we take $\sigma_1 = 100$, $\sigma_2 = 200$, $L = 80$ and sample f with the sampling rate $\Delta t = 1/10$. The

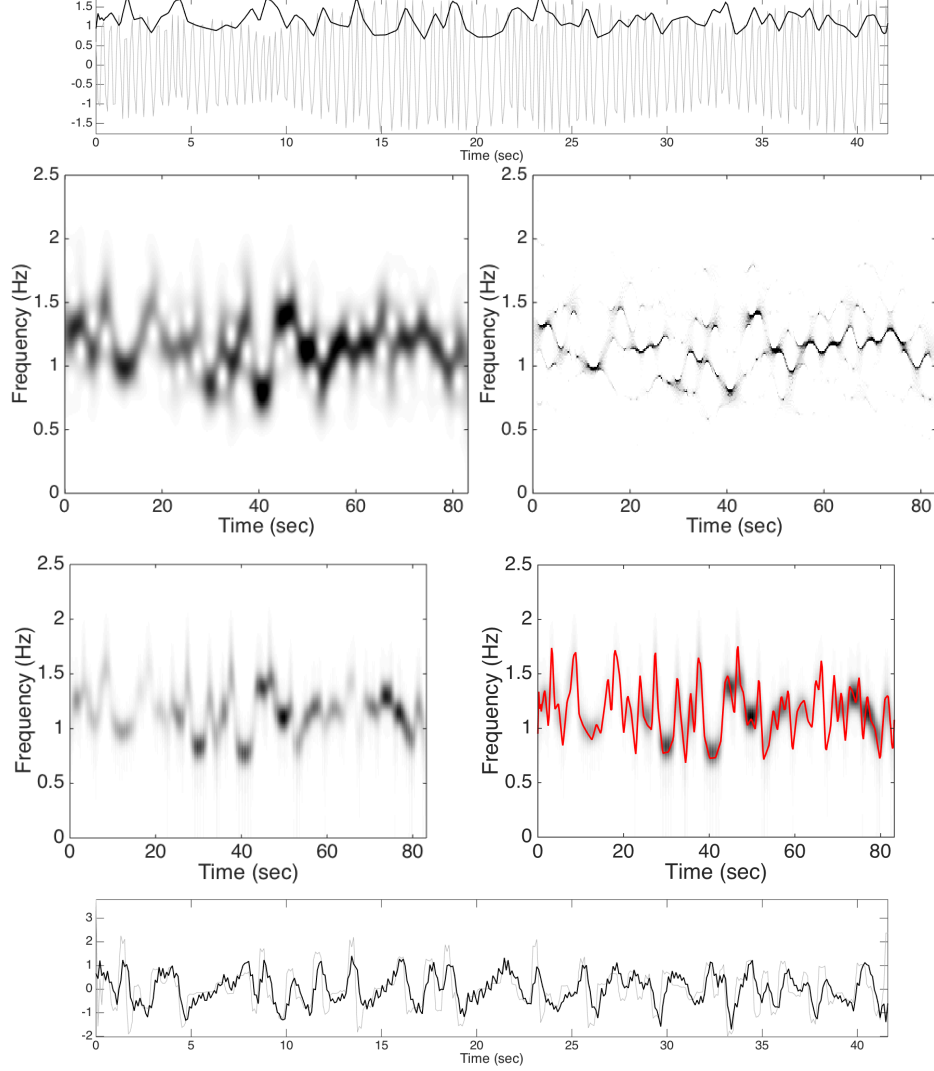


Figure 1: Top: the signal f_1 is shown as the gray curve with the instantaneous frequency superimposed as the black curve. It is clear that the instantaneous frequency varies fast. In the second row, the short time Fourier transform with the Gaussian window with the standard deviation 1 is shown on the left and the synchrosqueezed short time Fourier transform is shown on the right. In the third row, the Tycoon result is shown on the left and our result with the instantaneous frequency superimposed as a red curve is shown on the right. In the bottom, the chirp factor, $\phi_2''(t)$, is shown as the gray curve and the estimated $\phi_2''(t)$; that is, the $\alpha(t)$, is properly normalized and superimposed as the black curve. In the top and bottom figures, for the sake of visibility, only the first part of the signal is demonstrated.

result is shown in Figure 2. For the comparison purpose, we also show results from other TF analysis methods. In STFT and synchrosqueezed STFT, the window function is the same as that in the first example – the Gaussian window with the standard deviation

$\sigma = 1$. We also show the result with the synchrosqueezed CWT [15, 9], where the mother wavelet ψ is chosen to satisfy $\hat{\psi}(\xi) = e^{\frac{1}{(\frac{\xi-1}{0.2})^2-1}} \chi_{[0.8, 1.2]}$, where χ is the indicator function. Further, the popular empirical mode decomposition algorithm combined with the Hilbert spectrum (EMD-HS) [30] is also evaluated. The tvPS of f determined by EMD-HS is via the following steps. First, we run the proposed sifting process and decompose the given signal f into K components and the remainder term (see [30] for details of the sifting process); that is, $f(t) = \sum_{k=1}^{K_{\mathfrak{H}}} x_k(t) + r(t)$, where $K_{\mathfrak{H}} \in \mathbb{N}$ is chosen by the user, x_k is the k -th decomposed oscillatory component and r is the remainder term. The IF and AM of the k -th oscillatory component is determined by the Hilbert transform; that is, by $\tilde{x}_k(t) = x_k(t) + i\mathcal{H}(x_k(t)) = b_k(t)e^{i2\pi\psi_k(t)}$, where \mathcal{H} is the Hilbert transform, the IF and the AM of the k -th oscillatory component are estimated by $\psi'_k(t)$ and $b_k(t)$. Here we assume that x_k is well-behaved so that the Hilbert transform works. Finally, the tvPS (or called the Hilbert spectrum in the literature) of the signal f determined by the EMD-HS, denoted as \mathfrak{H}_f , is set to be $\mathfrak{H}_f(t, \omega) = \sum_{k=1}^{K_{\mathfrak{H}}} b_k(t)\delta(\omega - \psi'_k(t))$. In this work, due to the well-known mode-mixing issue of EMD and the number of components is not known a priori, we choose $K_{\mathfrak{H}} = 6$ so that we could hope to capture all needed information. We mention that one possible approach to evaluate the IF and AM after the sifting process is applying the SST directly to $x_k(t)$; this combination has been shown useful in the strong field atomic physics [45]. The results of STFT, synchrosqueezed STFT, synchrosqueezing CWT and EMD-HS are shown in Figure 3. Visually, it is clear that the proposed convex optimization approach, Tycoon, provides the dynamical information hidden inside the signal f , since the IFs of both components are better extracted in Tycoon, while several visually obvious artifacts could not be ignored in other TF analyses. For example, although we could see the overall pattern of the IF of f_2 in the STFT, the interfering pattern could not be ignored. While the IF of f_2 could be well captured in synchrosqueezed CWT, the IF of f_1 is blurred; on the other hand, while the IF of f_1 could be well captured in EMD-HS, the IF of f_2 is blurred. Clearly, the IF patterns of both components could not be easily identified in the synchrosqueezed STFT.

5.3. Performance quantification

To further quantify the performance of Tycoon, we consider the following metric. As indicated above, we would expect to recover the itvPS. Thus, to evaluate the performance of Tycoon and have a comparison with other TF analyses, we would compare the time varying power spectrum (tvPS) determined by different TF analyses with the itvPS of the clean simulated signal s . If we view both the itvPS and the tvPS as distributions on the TF-plane, we could apply the *Optimal Transport* (OT) distance, which is also well known as the Earth Mover distance (EMD), to evaluate how different the obtained tvPS is from the itvPS [17]. We would refer the reader to [49, section 2.2] for its detail theory. Here we quickly summarize how it works. Given two probability measures on the same set, the OT-distance evaluate the amount of “work” needed to “deform” one into the other. Precisely, the OT-distance between two probability distributions μ and ν on a metric space (\mathcal{S}, d) involves an optimization over all possible probability measures

on $\mathbf{S} \times \mathbf{S}$ that have μ and ν as marginals, denoted as $\mathcal{P}(\mu, \nu)$, by

$$d_{\text{OT}}(\mu, \nu) := \inf_{\rho \in \mathcal{P}(\mu, \nu)} \int d(x, y) d\rho(x, y) , \quad (60)$$

which in the one-dimensional case, that is, when $\mathbf{S} \subset \mathbb{R}$, and d is the canonical Euclidean distance, $d(x, y) = |x - y|$, could be easily evaluated. Define $f_\mu(x) = \int_{-\infty}^x d\mu$ and $f_\nu(x) = \int_{-\infty}^x d\nu$, the OT distance is reduced to the L^1 difference of f_μ and f_ν ; that is,

$$d_{\text{OT}}(\mu, \nu) = \int_{\mathbf{S}} |f_\mu(x) - f_\nu(x)| dx . \quad (61)$$

In the TF representation, as tvPS is always non-negative, we could view the distribution of the tvPS at each time as a probability density after normalizing its L^1 to 1. This distribution indicates how accurate the TF analyses recover the oscillatory behavior of the signal at each time. Thus, based on the OT distance, we consider the following \mathfrak{D} metric to evaluate the performance of each TF analyses of the function f by

$$\mathfrak{D} := 100 \times \int_{-\infty}^{\infty} d_{\text{OT}}(P_f^t, \tilde{P}_f^t) dt, \quad (62)$$

where $P_f^t(\omega) := \frac{S_f(t, \omega)}{\int_0^\infty S_f(t, \eta) d\eta}$, $\tilde{P}_f^t(\omega) := \frac{\tilde{S}_f(t, \omega)}{\int_0^\infty \tilde{S}_f(t, \eta) d\eta}$, $S_f(t, \omega)$ is the itvPS and $\tilde{S}_f(t, \omega)$ is the estimated tvPS by a chosen TF analysis. Clearly, the small the \mathfrak{D} metric is, the better the itvPS is approximated.

To evaluate the second example, we run STFT, synchrosqueezed STFT, synchrosqueezing CWT and Tycoon on 100 different realizations of f_2 in (59), and evaluate the \mathfrak{D} metric. The result is displayed in (mean \pm standard deviation). The \mathfrak{D} metric between the itvPS and the tvPS determined by Tycoon (respectively, EMD-HS, STFT, synchrosqueezed STFT and synchrosqueezed CWT) is 6.06 ± 0.25 (respectively, 7.18 ± 0.93 , 8.76 ± 0.41 , 8.13 ± 0.42 and 7.36 ± 0.67). Further, under the null hypothesis that there is no performance difference between the tvPS determined by Tycoon and STFT evaluated by the \mathfrak{D} metric and we set the significant level at 5%, the t-test rejects the null hypothesis with the p-value less than 10^{-8} . The same hypothesis testing results hold for the comparison between Tycoon and other methods. Note that while the performance of Tycoon seems better than EMD-HS, the \mathfrak{D} metric only reflects partial information regarding the difference and more details should be taken into account to achieve a fair comparison. For example, if we set $K_{\mathfrak{S}} = 2$, the \mathfrak{D} metric between the itvPS and the tvPS determined by EMD-HS becomes 4.98 ± 0.81 , which might suggest that EMD-HS performs better. However, this “better performance” is not surprising since the sparsity property is perfectly satisfies in EMD-HS, which is inherited in the procedure, while the mode mixing issue might lead to wrong interpretation eventually. Note that it is also possible to post-process the outcome of the sifting process to enhance the result, but these ad-hoc post-processing again are not mathematically well supported. Since it is out of the scope of this paper, we would leave this overall comparison between different TF analyses based on different philosophy as well as a better metric to the future work.

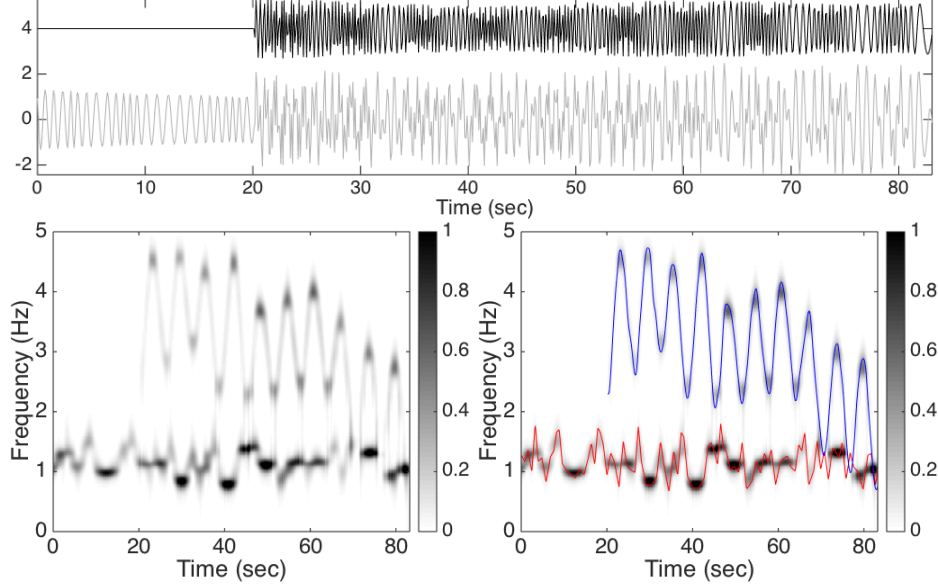


Figure 2: Top: the signal f is shown as the gray curve with f_2 superimposed as the black curve which is shifted up by 4 to increase the visualization. It is clear that the instantaneous frequency (IF) also varies fast in both components. In the bottom row, the intensity of the time frequency representation, $|\hat{R}_f|^2$, determined by the proposed Tycoon algorithm is shown on the left; on the right hand side, the instantaneous frequencies associated with the two components are superimposed on $|\hat{R}_f|^2$ as a red curve and a blue curve.

5.4. Two component, noisy

In the third example, we add noise to the signal f and see how the proposed algorithm performs. To model the noise, we define the signal to noise ratio (SNR) as

$$\text{SNR} := 20 \log_{10} \frac{\text{std}(f)}{\text{std}(\Phi)}, \quad (63)$$

where f is the clean signal, Φ is the added noise and std means the standard deviation. In this simulation, we add the Gaussian white noise with SNR 7.25 to the clean signal f , and obtain a noisy signal Y . The result is shown in Figure 4. Clearly, we see that even when noise exists, the algorithm provides a reasonable result. To further evaluate the performance, we run STFT, synchrosqueezed STFT, synchrosqueezing CWT and Tycoon on 100 different realizations of f_2 in (59) as well as 100 different realizations of noise, and evaluate the \mathfrak{D} metric. Here we use the same parameters as those in the second example to run STFT, synchrosqueezed STFT and synchrosqueezed CWT. Since it is well known that EMD is not robust to noise, we replace the sifting process in EMD by that of the ensemble EMD (EEMD) to decompose the signal into $K_{\mathfrak{J}} = 6$ oscillatory components, and generate the tvPS by the Hilbert transform as that in EMD. We call the method EEMD-HS. See [51] for the detail of the EEMD algorithm. The \mathfrak{D} metric between the itvPS and the tvPS determined by Tycoon (respectively, EEMD-HS, STFT, synchrosqueezed STFT and synchrosqueezed CWT) is 11.87 ± 0.74 (respectively,

11.65 ± 0.63 , 14.53 ± 0.55 , 14.09 ± 0.58 and 12.79 ± 0.69). The same hypothesis testing shows the significant difference between the performance of Tycoon and that of STFT, synchrosqueezed STFT and synchrosqueezed CWT, while there is no significant difference between the performance of Tycoon and that of EEMD-HS. Again, the same comments for the comparison between Tycoon and EMD-HS carry here when we compare Tycoon and EEMD-HS, and we leave the details to the future work.

6. Discussion and future work

In this paper we propose a generalized intrinsic mode functions and adaptive harmonic model to model oscillatory functions with fast varying instantaneous frequency. A convex optimization approach to find the time-frequency representation, referred to as Tycoon algorithm, is proposed. While the numerical results are encouraging, there are several things we should discuss.

1. While with the help of FISTA the optimization process can be carried out, it is still not numerically efficient enough for practical usage. For example, it takes about 3 minutes to finish analyzing a time series with 512 points in the laptop, but in many problems the data length is of order 10^5 or longer. Finding a more efficient strategy to carry out the optimization is an important future work. One possible solution is by the sliding window idea. For a given long time series f of length n and a length $m < n$, we could run the optimization consecutively on the subinterval $I_j := [j - m, j + m]$ to determine the tvPS at time j . Thus, the overall computational complexity could be $O(F(m)n)$, where $F(m)$ is the complexity of running the optimization on the subinterval I_j .
2. When there are more than one oscillatory component, we could consider (27) to improve the result. However, in practice it does not significantly improve the result. Since it is of its own interest, we decide to leave it to the future work.
3. While the Tycoon algorithm is not very much sensitive to the choice of parameters μ , λ and γ , how to choose an optimal set of parameters is left unanswered in the current paper.
4. The noise behavior and influence on the Tycoon algorithm is not clear at this moment, although we could see that it is robust to the existence of noise in the numerical section. Theoretically studying the noise influence on the algorithm is important for us to better understand what we see in practice.

Before closing the paper, we would like to indicate an interesting finding about SST which is related to our current study. When an oscillatory signal is composed of intrinsic mode type function with slowly varying IF, it has been studied that the time-frequency representation of a function depends “weakly” on a chosen window, when the window has a small support in the Fourier domain [15, 9]. Precisely, the result depends only on the first three absolute moments of the chosen window and its derivative, but not depends on the profile of the window itself. However, the situation is different when we consider an oscillatory signal composed of gIMT function with fast varying IF. As we have shown in Figure 2, when the window is chosen to have a small support in the Fourier domain, the STFT and synchrosqueezed STFT results are not ideal. Nevertheless, nothing prevents us from trying a window with a small support in the time domain; that is, a wide support

in the Fourier domain. As is shown in Figure 5, by taking the window to be a Gaussian function with the standard deviation 0.4, STFT and synchrosqueezed STFT provide reasonable results for the signal f considered in (59). Note that while we could start to see the dynamics in both STFT and synchrosqueezed STFT, the overall performance is not as good as that provided by Tycoon. Since it is not the focus of the current paper, we just indicate the possibility of achieving a better time-frequency representation by choosing a suitable window in SST, but not make effort to determine the optimal window. This kind of approach has been applied to the strong field atomic physics [35, 45], where the window is manually but carefully chosen to extract the physically meaningful dynamics. A theoretical study regarding this topic will be reported in the near future.

7. Acknowledgement

Hau-tieng Wu would like to thank Professor Ingrid Daubechies and Professor Andrey Feuerverger for their valuable discussion and Dr. Su Li for discussing the application direction in music and sound analyses. Hau-tieng Wu's work is partially supported by Sloan Research Fellow FR-2015-65363. Part of this work was done during Hau-tieng Wu's visit to National Center for Theoretical Sciences, Taiwan, and he would like to thank NCTS for its hospitality. Matthieu Kowalski benefited from the support of the "FMJH Program Gaspard Monge in optimization and operation research", and from the support to this program from EDF. We would also like to thank the anonymous reviewers for their constructive and helpful comments.

References

- [1] F. Auger, E. Chassande-Mottin, and P. Flandrin. Making reassignment adjustable: The levenberg-marquardt approach. In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, pages 3889–3892, March 2012.
- [2] F. Auger and P. Flandrin. Improving the readability of time-frequency and time-scale representations by the reassignment method. *IEEE Trans. Signal Process.*, 43(5):1068–1089, may 1995.
- [3] P. Balazs, M. Dörfler, F. Jaillet, N. Holighaus, and G. Velasco. Theory, implementation and applications of nonstationary Gabor frames. *Journal of Computational and Applied Mathematics*, 236(6):1481–1496, 2011.
- [4] A. Beck and M. Teboulle. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *Image Processing, IEEE Transactions on*, 18(11):2419–2434, 2009.
- [5] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sciences*, 2(1):183–202, 2009.
- [6] Jérôme Bolte, Shoham Sabach, and Marc Teboulle. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Mathematical Programming*, 146(1-2):459–494, 2014.
- [7] A. Chambolle and C. Dossal. On the convergence of the iterates of FISTA. *Preprint hal-01060130*, September, 2014.
- [8] E. Chassande-Mottin, F. Auger, and P. Flandrin. Time-frequency/time-scale reassignment. In *Wavelets and signal processing*, Appl. Numer. Harmon. Anal., pages 233–267. Birkhäuser Boston, Boston, MA, 2003.
- [9] Y.-C. Chen, M.-Y. Cheng, and H.-T. Wu. Nonparametric and adaptive modeling of dynamic seasonality and trend with heteroscedastic and dependent errors. *J. Roy. Stat. Soc. B*, 76:651–682, 2014.
- [10] C. K. Chui, Y.-T. Lin, and H.-T. Wu. Real-time dynamics acquisition from irregular samples – with application to anesthesia evaluation. *Analysis and Applications, accepted for publication*, 2015. DOI: 10.1142/S0219530515500165.

- [11] C. K. Chui and H.N. Mhaskar. Signal decomposition and analysis via extraction of frequencies. *Appl. Comput. Harmon. Anal.*, 2015.
- [12] A. Cicone, J. Liu, and H. Zhou. Adaptive local iterative filtering for signal decomposition and instantaneous frequency analysis. *arXiv preprint arXiv:1411.6051*, 2014.
- [13] A. Cicone and H. Zhou. Multidimensional iterative filtering method for the decomposition of high-dimensional non-stationary signals. *arXiv preprint arXiv:1507.07173*, 2015.
- [14] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Modeling & Simulation*, 4(4):1168–1200, 2005.
- [15] I. Daubechies, J. Lu, and H.-T. Wu. Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool. *Appl. Comput. Harmon. Anal.*, 30:243–261, 2011.
- [16] I. Daubechies and S. Maes. A nonlinear squeezing of the continuous wavelet transform based on auditory nerve models. *Wavelets in Medicine and Biology*, pages 527–546, 1996.
- [17] I. Daubechies, Y. Wang, and H.-T. Wu. ConceFT: Concentration of frequency and time via a multitapered synchrosqueezing transform. *Philosophical Transactions A*, Accepted for publication, 2015.
- [18] A. M. De Livera, R. J. Hyndman, and R. D. Snyder. Forecasting Time Series With Complex Seasonal Patterns Using Exponential Smoothing. *J. Am. Stat. Assoc.*, 106(496):1513–1527, 2011.
- [19] C. Deledalle, S. Vaiter, G. Peyré, J. Fadili, and C. Dossal. Proximal splitting derivatives for risk estimation. *Journal of Physics: Conference Series*, 386(1):012003, 2012.
- [20] K. Dragomiretskiy and D. Zosso. Variational Mode Decomposition. *IEEE Trans. Signal Process.*, 62(2):531–544, 2014.
- [21] P. Flandrin. *Time-frequency/time-scale analysis*, volume 10 of *Wavelet Analysis and its Applications*. Academic Press Inc., 1999.
- [22] P. Flandrin. Time frequency and chirps. In *Proc. SPIE*, volume 4391, pages 161–175, 2001.
- [23] G. Galiano and J. Velasco. On a non-local spectrogram for denoising one-dimensional signals. *Applied Mathematics and Computation*, 244:1–13, 2014.
- [24] J. Gilles. Empirical Wavelet Transform. *IEEE Trans. Signal Process.*, 61(16):3999–4010, 2013.
- [25] E. T. Hale, W. Yin, and Y. Zhang. Fixed-point continuation for ℓ_1 -minimization: Methodology and convergence. *SIAM Journal on Optimization*, 19(3):1107–1130, 2008.
- [26] T. Hou and Z. Shi. Data-driven time-frequency analysis. *Appl. Comput. Harmon. Anal.*, 35(2):284 – 308, 2013.
- [27] T. Hou and Z. Shi. Sparse time-frequency representation of nonlinear and nonstationary data. *Science China Mathematics*, 56(12):2489–2506, 2013.
- [28] T. Y. Hou and Z. Shi. Adaptive data analysis via sparse time-frequency representation. *Adv. Adapt. Data Anal.*, 03(01n02):1–28, 2011.
- [29] C. Huang, Y. Wang, and L. Yang. Convergence of a convolution-filtering-based algorithm for empirical mode decomposition. *Adv. Adapt. Data Anal.*, 1(4):561–571, 2009.
- [30] N. E. Huang, Z. Shen, S. R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Lond. A*, 454(1971):903–995, 1998.
- [31] Z. Huang, J. Zhang, T. Zhao, and Y. Sun. Synchrosqueezing s-transform and its application in seismic spectral decomposition. *Geoscience and Remote Sensing, IEEE Transactions on*, PP(99):1–9, 2015.
- [32] A. Jahangir, Lee V., P.A. Friedman, J.M. Trusty, D.O. Hodge, and et al. Long-term progression and outcomes with aging in patients with lone atrial fibrillation: a 30-year follow-up study. *Circulation*, 115:3050–3056, 2007.
- [33] K. Kodera, R. Gendrin, and C. Villedary. Analysis of time-varying signals with small bt values. *IEEE Trans. Acoust., Speech, Signal Processing*, 26(1):64 – 76, feb 1978.
- [34] C. Li and M. Liang. A generalized synchrosqueezing transform for enhancing signal time-frequency representation. *Signal Processing*, 92(9):2264 – 2274, 2012.
- [35] P.-C. Li, Y.-L. Sheu, C. Laughlin, and S.-I. Chu. Dynamical origin of near- and below-threshold harmonic generation of Cs in an intense mid-infrared laser field. *Nature Communication*, 6, 2015.
- [36] L. Lin, Y. Wang, and H. Zhou. Iterative filtering as an alternative for empirical mode decomposition. *Adv. Adapt. Data Anal.*, 1(4):543–560, 2009.
- [37] C. Liu, T. Y. Hou, and Z. Shi. On the uniqueness of sparse time-frequency representation of multiscale data. *Multiscale Model. and Simul.*, 13(3):790–811, 2015.
- [38] Ignace Loris. On the performance of algorithms for the minimization of 1-penalized functionals. *Inverse Problems*, 25(3):035008, 2009.
- [39] S. Mann and S. Haykin. The chirplet transform: physical considerations. *Signal Process. IEEE*

- Trans.*, 43(11):2745–2761, 1995.
- [40] V. A. Morozov. On the solution of functional equations by the method of regularization. *Soviet Math. Dokl.*, 7(1):414–417, 1966.
 - [41] T. Oberlin, S. Meignen, and V. Perrier. An alternative formulation for the Empirical Mode Decomposition. *IEEE Trans. Signal Process.*, 60(5):2236–2246, 2012.
 - [42] T. Oberlin, S. Meignen, and V. Perrier. Second-order synchrosqueezing transform or invertible reassignment? towards ideal time-frequency representations. *IEEE Trans. Signal Process.*, 63(5):1335–1344, March 2015.
 - [43] N. Pustelnik, P. Borgnat, and P. Flandrin. Empirical mode decomposition revisited by multicomponent non-smooth convex optimization. *Signal Processing*, 102(0):313 – 331, 2014.
 - [44] B. Ricaud, G. Stempfel, and B. Torrèsani. An optimally concentrated Gabor transform for localized time-frequency components. *Adv Comput Math*, 40:683–702, 2014.
 - [45] Y.-L. Sheu, H.-T. Wu, and L.-Y. Hsu. Exploring laser-driven quantum phenomena from a time-frequency analysis perspective: A comprehensive study. *Optics Express*, 23:30459–30482, 2015.
 - [46] R.G. Stockwell, L. Mansinha, and R.P. Lowe. Localization of the complex spectrum: the S transform. *Signal Process. IEEE Trans.*, 44(4):998–1001, 1996.
 - [47] P. Tavallali, T. Hou, and Z. Shi. Extraction of intrawave signals using the sparse time-frequency representation method. *Multiscale Modeling & Simulation*, 12(4):1458–1493, 2014.
 - [48] P. Tseng. Approximation accuracy, gradient methods, and error bound for structured convex optimization. *Mathematical Programming*, 125(2):263–295, 2010.
 - [49] C. Villani. *Topics in Optimal Transportation*. Graduate Studies in Mathematics, American Mathematical Society, 2003.
 - [50] H.-T. Wu. Instantaneous frequency and wave shape functions (I). *Appl. Comput. Harmon. Anal.*, 35:181–199, 2013.
 - [51] Z. Wu and N. E. Huang. Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Adv. Adapt. Data Anal.*, 1:1 – 41, 2009.
 - [52] H. Yang. Synchrosqueezed Wave Packet Transforms and Diffeomorphism Based Spectral Analysis for 1D General Mode Decompositions. *Appl. Comput. Harmon. Anal.*, 39:33–66, 2014.
 - [53] W. I. Zangwill. *Nonlinear programming: a unified approach*, volume 196. Prentice-Hall Englewood Cliffs, NJ, 1969.

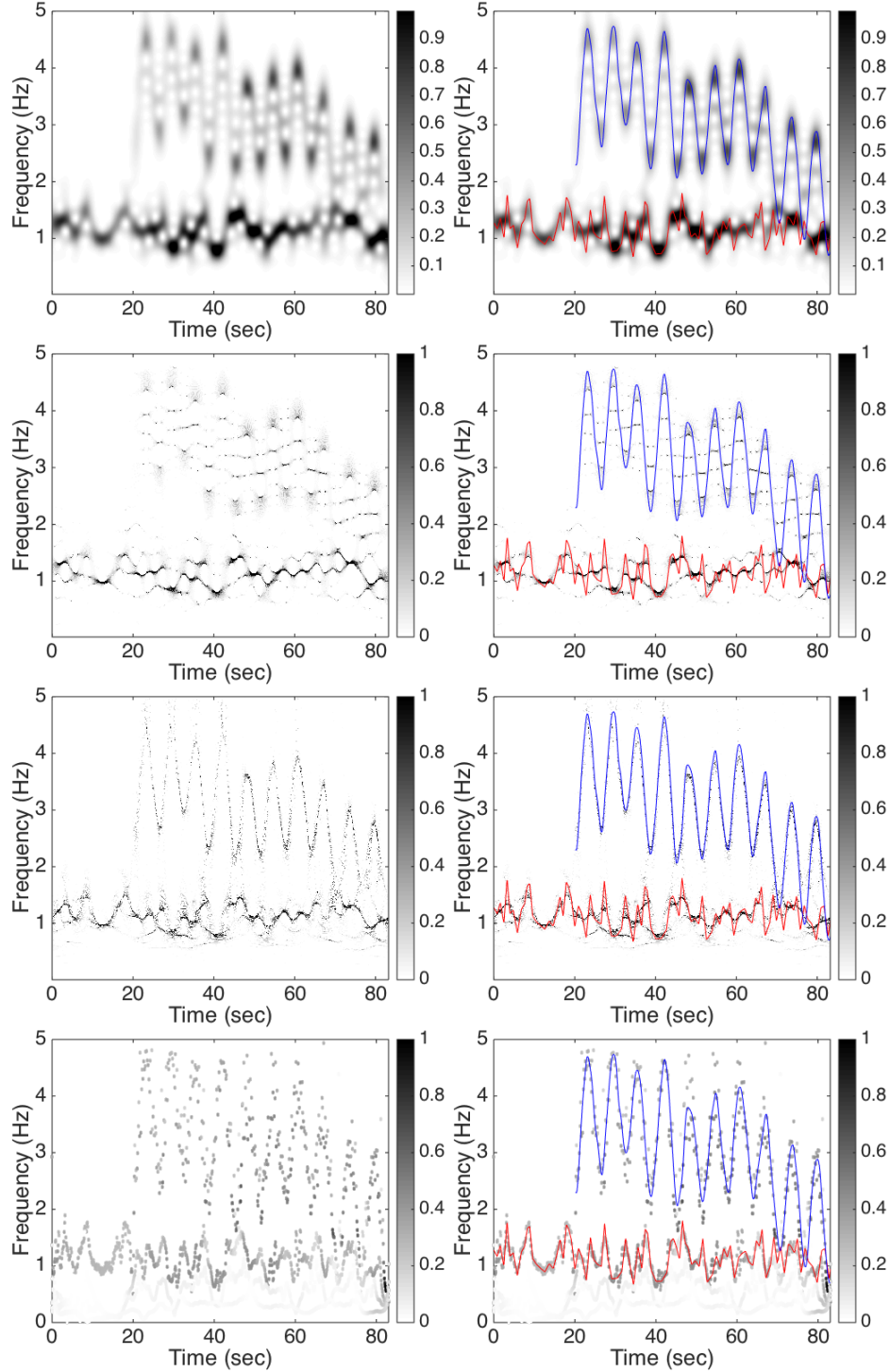


Figure 3: The time frequency (TF) representations²⁷ of different TF analyses on the signal f . In the first row, on the left, the short time Fourier transform (STFT) with a Gaussian window with the standard deviation $\sigma = 1$ is shown, and on the right the IF's of both components are superimposed for the visual comparison. In the second row, on the left, the synchrosqueezed STFT with a Gaussian window with the standard deviation $\sigma = 1$ is shown, and on the right the IF's of both components are superimposed for the visual comparison. In the third row, on the left, we show the synchrosqueezed continuous wavelet

transform with the mother wavelet ψ so that $\hat{\psi}(\xi) = e^{\frac{1}{(\frac{\xi-1}{0.2})^2 - 1}} \chi_{[0.8, 1.2]}$, where χ is the indicator function, and on the right the IF's of both components are superimposed for the visual comparison. It is clear that the slowly oscillatory component is not well captured. In the bottom row, on the left, we show the TF representation determined by the empirical mode decomposition with the Hilbert transform, and on the right the IF's of both components are superimposed for the visual inspection. It is clear that the fast oscillatory component is not well captured.

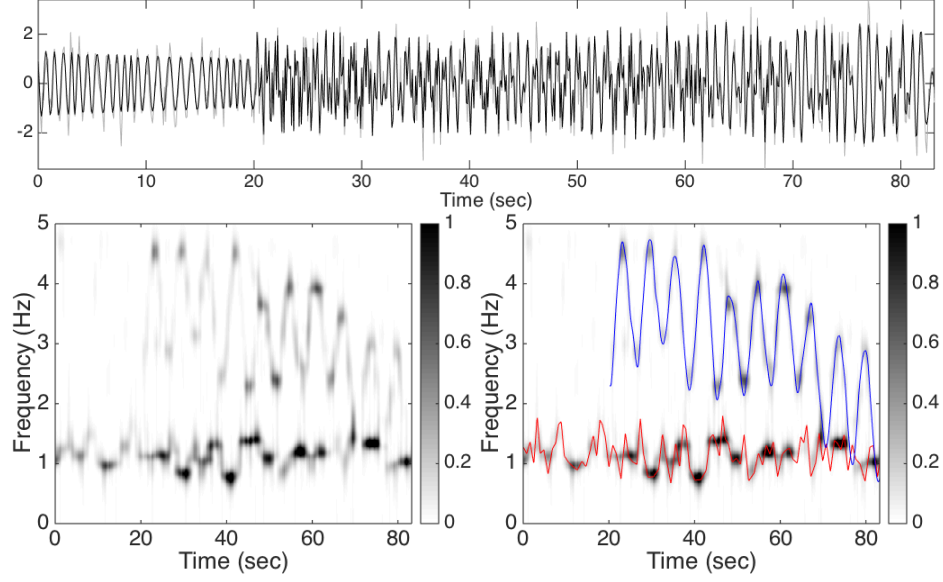


Figure 4: Top: the noisy signal Y is shown as the gray curve with the clean signal f superimposed as the black curve. In the second row, the intensity of the time frequency representation, $|\hat{R}_Y|^2$, determined by our proposed Tycoon algorithm is shown on the left; on the right hand side, the instantaneous frequencies associated with the two components are superimposed on $|\hat{R}_Y|^2$ as a red curve and a blue curve.

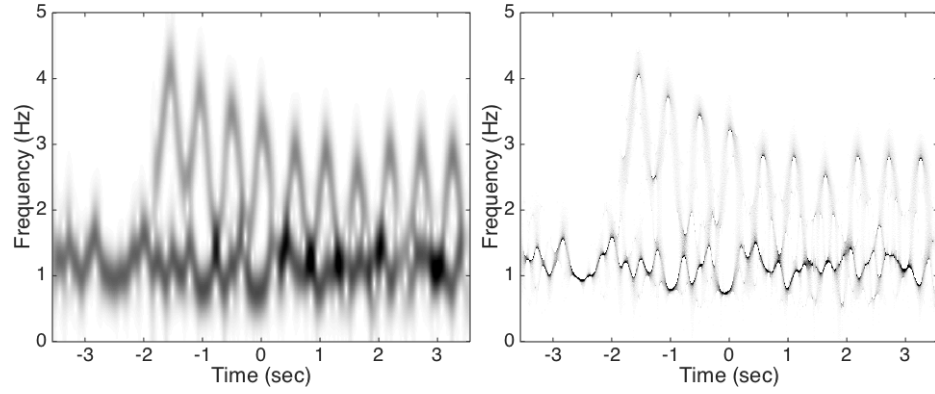


Figure 5: Left: the intensity of the short time Fourier transform (STFT) with a Gaussian window with the standard deviation $\sigma = 0.4$ is shown on the left and the intensity of the synchrosqueezed STFT is shown on the right.

Appendix A. Proof of Theorem 2.1

Suppose

$$g(t) = a(t) \cos \phi(t) = (a(t) + \alpha(t)) \cos(\phi(t) + \beta(t)) \in \mathcal{Q}_\epsilon^{c_1, c_2, c_3}. \quad (\text{A.1})$$

Clearly we know $\alpha \in C^1(\mathbb{R})$, $\beta \in C^3(\mathbb{R})$. By the definition of $\mathcal{Q}_\epsilon^{c_1, c_2, c_3}$, we have

$$\inf_{t \in \mathbb{R}} a(t) > c_1, \quad \sup_{t \in \mathbb{R}} a(t) < c_2, \quad (\text{A.2})$$

$$\inf_{t \in \mathbb{R}} \phi'(t) > c_1, \quad \sup_{t \in \mathbb{R}} \phi'(t) < c_2, \quad |\phi''(t)| \leq c_3 \quad (\text{A.3})$$

$$|a'(t)| \leq \epsilon \phi'(t), \quad |\phi'''(t)| \leq \epsilon \phi'(t) \quad (\text{A.4})$$

and

$$\inf_{t \in \mathbb{R}} [a(t) + \alpha(t)] > c_1, \quad \sup_{t \in \mathbb{R}} [a(t) + \alpha(t)] < c_2, \quad (\text{A.5})$$

$$\inf_{t \in \mathbb{R}} [\phi'(t) + \beta'(t)] > c_1, \quad \sup_{t \in \mathbb{R}} [\phi'(t) + \beta'(t)] < c_2, \quad |\phi''(t) + \beta''(t)| \leq c_3 \quad (\text{A.6})$$

$$|a'(t) + \alpha'(t)| \leq \epsilon(\phi'(t) + \beta'(t)), \quad |\phi'''(t) + \beta'''(t)| \leq \epsilon(\phi'(t) + \beta'(t)). \quad (\text{A.7})$$

The proof is divided into two parts. The first part is determining the restrictions on the possible β and α based on the positivity condition of $\phi'(t)$ and $a(t)$, which is independent of the conditions (A.4) and (A.7). The second part is to control the amplitude of β and α , which depends on the conditions (A.4) and (A.7).

First, based on the conditions (A.2), (A.3), (A.5) and (A.6), we show how β and α are restricted. By the monotonicity of $\phi(t)$ based on the condition (A.3), define $t_m \in \mathbb{R}$, $m \in \mathbb{Z}$, so that $\phi(t_m) = (m + 1/2)\pi$ and $s_m \in \mathbb{R}$, $m \in \mathbb{Z}$, so that $\phi(s_m) = m\pi$. In other words, we have

$$g(t_m) = 0 \quad \text{and} \quad g(s_m) = (-1)^m a(s_m).$$

Thus, for any $n \in \mathbb{Z}$, when $t = t_n$, we have

$$\begin{aligned} & (a(t_n) + \alpha(t_n)) \cos(\phi(t_n) + \beta(t_n)) \\ &= (a(t_n) + \alpha(t_n)) \cos[n\pi + \pi/2 + \beta(t_n)] \\ &= a(t_n) \cos(n\pi + \pi/2) = 0, \end{aligned} \quad (\text{A.8})$$

where the second equality comes from (A.1). This leads to $\beta(t_n) = k_n\pi$, $k_n \in \mathbb{Z}$, since $a(t_n) + \alpha(t_n) > 0$ by (A.6).

Lemma 1. *k_n are the same for all $n \in \mathbb{Z}$ and k_n are even. As changing the phase function globally by $2l\pi$, where $l \in \mathbb{Z}$, will not change the value of $g(t_n)$ for all $n \in \mathbb{Z}$, we could assume that $\beta(t_m) = 0$ for all $m \in \mathbb{Z}$.*

Proof. Suppose there exists t_n so that $\beta(t_n) = k\pi$ and $\beta(t_{n+1}) = (k+l)\pi$, where $k, l \in \mathbb{Z}$ and $l > 0$. In other words, we have $\phi(t_{n+1}) = \phi(t_n) + (l+1)\pi$. By the smoothness of β , we know there exists at least one $t' \in (t_n, t_{n+1})$ so that $\phi(t') + \beta(t') = (n+3/2)\pi$, but this

is absurd since it means that $(a(t) + \alpha(t))\cos(\phi(t) + \beta(t))$ will change sign in (t_n, t_{n+1}) while $a(t)\cos(\phi(t))$ will not.

Suppose k_n is a fixed odd integer k , then since $\beta \in C^3(\mathbb{R})$ and $\beta(t_n) = \beta(t_{n+1}) = k\pi$, there exists $t' \in (t_n, t_{n+1})$ so that $\beta(t') = k\pi$ and hence

$$a(t')\cos(\phi(t')) = (a(t') + \alpha(t'))\cos(\phi(t') + \beta(t')) = -(a(t') + \alpha(t'))\cos(\phi(t')),$$

which is again absurd since $\cos(\phi(t')) \neq 0$ and the amplitudes are positive by (A.2) and (A.5). We thus obtain the second claim. \square

Lemma 2. $\beta'(t)$ is 0 or changes sign inside $[t_n, t_{n+1}]$ for all $n \in \mathbb{Z}$. Furthermore, $|\beta(t') - \beta(t'')| < \pi$ for any $t', t'' \in [t_m, t_{m+1}]$ for all $m \in \mathbb{Z}$.

Proof. By the fundamental theorem of calculus and the fact that $\beta(t_n) = \beta(t_{n+1}) = 0$, we know that

$$0 = \beta(t_{n+1}) - \beta(t_n) = \int_{t_n}^{t_{n+1}} \beta'(u) du.$$

which implies the first argument. Also, due to the monotonicity of $\phi + \beta$ (A.6), that is, $(n + 1/2)\pi = \phi(t_n) + \beta(t_n) < \phi(t') + \beta(t') < \phi(t_{n+1}) + \beta(t_{n+1}) = (n + 3/2)\pi$ for all $t' \in (t_n, t_{n+1})$, we have the second claim

$$|\beta(t') - \beta(t'')| < \pi.$$

Indeed, if $|\beta(t') - \beta(t'')| \geq \pi$, for some $t', t'' \in [t_n, t_{n+1}]$ and $t' < t''$, we get a contradiction since $\phi(t'') + \beta(t'') \notin [(n + 1/2)\pi, (n + 3/2)\pi]$ while $\phi(t') + \beta(t') \in [(n + 1/2)\pi, (n + 3/2)\pi]$. \square

Lemma 3. $\frac{a(s_n)}{a(s_n) + \alpha(s_n)} = \cos(\beta(s_n))$ for all $n \in \mathbb{Z}$. In particular, $\alpha(s_m) = 0$ if and only if $\beta(s_m) = 0$, $m \in \mathbb{Z}$.

Proof. When $t = s_m$, we have

$$\begin{aligned} (-1)^m a(s_m) &= a(s_m) \cos(m\pi) \\ &= (a(s_m) + \alpha(s_m)) \cos[m\pi + \beta(s_m)] \\ &= (-1)^m (a(s_m) + \alpha(s_m)) \cos(\beta(s_m)), \end{aligned} \tag{A.9}$$

where the second equality comes from (A.1), which leads to $\alpha(s_m) \geq 0$ since $|\cos(\beta(s_m))| \leq 1$.

Notice that (A.9) implies that $\beta(s_m) = 2k_m\pi$, where $k_m \in \mathbb{Z}$, if and only if $\alpha(s_m) = 0$. Without loss of generality, assume $k_m > 0$. Since $\beta \in C^3(\mathbb{R})$, there exists $t' \in (t_{m-1}, s_m)$ so that $\beta(t') = \pi$ and hence

$$a(t')\cos(\phi(t')) = (a(t') + \alpha(t'))\cos(\phi(t') + \beta(t')) = -(a(t') + \alpha(t'))\cos(\phi(t')),$$

which is absurd since $\cos(\phi(t')) \neq 0$ and the positive amplitudes by (A.2) and (A.5). Thus we conclude that $\beta(s_m) = 0$.

To show the last part, note that when $\alpha(s_m) > 0$, $0 < \cos(\beta(s_m)) = \frac{a(s_m)}{a(s_m) + \alpha(s_m)} < 1$ by (A.9). Thus, we know $\beta(s_m) \in (-\pi/2, \pi/2) + 2n_m\pi$, where $n_m \in \mathbb{Z}$. By the same

argument as in the above, if $n_m > 0$, there exists $t' \in (t_{m-1}, s_m)$ so that $\beta(t') = \pi$ and hence

$$a(t') \cos(\phi(t')) = (a(t') + \alpha(t')) \cos(\phi(t') + \beta(t')) = -(a(t') + \alpha(t')) \cos(\phi(t')),$$

which is absurd since $\cos(\phi(t')) \neq 0$ and the positive amplitudes by (A.2) and (A.5). \square

Lemma 4. $\frac{a(t_n)}{a(t_n) + \alpha(t_n)} = \frac{\phi'(t_n) + \beta'(t_n)}{\phi'(t_n)}$ for all $n \in \mathbb{Z}$. In particular, $\alpha(t_n) = 0$ if and only if $\beta'(t_n) = 0$, $n \in \mathbb{Z}$.

Proof. For $0 < x \ll 1$, we have

$$(a(t_n + x) + \alpha(t_n + x)) \cos(\phi(t_n + x) + \beta(t_n + x)) = a(t_n + x) \cos(\phi(t_n + x)),$$

which means that

$$\frac{a(t_n + x)}{a(t_n + x) + \alpha(t_n + x)} = \frac{\cos(\phi(t_n + x) + \beta(t_n + x))}{\cos(\phi(t_n + x))}.$$

By the smoothness of ϕ and β , as $x \rightarrow 0$, the right hand side becomes

$$\begin{aligned} & \lim_{x \rightarrow 0} \frac{\cos(\phi(t_n + x) + \beta(t_n + x))}{\cos(\phi(t_n + x))} \\ &= \lim_{x \rightarrow 0} \frac{(\phi'(t_n + x) + \beta'(t_n + x)) \sin(\phi(t_n + x) + \beta(t_n + x))}{\phi'(t_n + x) \sin(\phi(t_n + x))} \\ &= \frac{\phi'(t_n) + \beta'(t_n)}{\phi'(t_n)}. \end{aligned}$$

Thus, since $a(t_n + x) + \alpha(t_n + x) > 0$ and $a(t_n + x) > 0$ for all x , we have

$$\frac{a(t_n)}{a(t_n) + \alpha(t_n)} = \frac{\phi'(t_n) + \beta'(t_n)}{\phi'(t_n)}.$$

\square

Lemma 5. $\beta''(t)$ is 0 or changes sign inside $[t_n, t_{n+1}]$ for all $n \in \mathbb{Z}$.

Proof. This is clear since $\beta'(t)$ is 0 or changes sign inside $[t_n, t_{n+1}]$ for all $n \in \mathbb{Z}$ by Lemma 2. \square

In summary, while $\beta(t_m) = 0$ for all $m \in \mathbb{Z}$, in general we loss the control of α at t_m . On s_m , α is directly related to β by Lemma 3; on t_m , α is directly related to β' by Lemma 4. We could thus call t_m and s_m the *hinging points* associated with the function g . Note that the control of α and β on the hinging points does not depend on β'' 's condition.

To finish the second part of the proof, we have to consider the conditions (A.4) and (A.7).

Lemma 6. $|\alpha(t)| \leq 2\pi\epsilon$ for all $t \in \mathbb{R}$. Further, we have $|\beta'(t_n)| \leq \frac{4\pi\phi'(t_n)}{a(t_n)}\epsilon$ for all $n \in \mathbb{Z}$.

Proof. Suppose there exists t' so that $\alpha(t') > 2\pi\epsilon$. The case $\alpha(t') < -2\pi\epsilon$ can be proved in the same way. Take $m \in \mathbb{Z}$ so that $t' \in (t_m, t_{m+1}]$. From (A.4) and (A.7) we have

$$|\alpha'(t)| \leq \epsilon(2\phi'(t) + \beta'(t)).$$

Thus, take $t \in (t_m, t_{m+1})$. Without loss of generality, we could assume $t \in (t_m, t')$, we have by the fundamental theorem of calculus

$$\begin{aligned} |\alpha(t') - \alpha(t)| &\leq \int_t^{t'} |\alpha'(u)| du \leq \epsilon[2\phi(t') - 2\phi(t) + \beta(t') - \beta(t)] \\ &\leq \epsilon[(\phi(t_{m+1}) + \beta(t_{m+1}) - \phi(t_m) - \beta(t_m)) + (\phi(t_{m+1}) - \phi(t_m))] \leq 2\pi\epsilon, \end{aligned}$$

where the last inequality holds due to the fact that $\phi + \beta$ and ϕ are both monotonic and Lemma 2. This fact leads to $\alpha(t) > 0$ for all $t \in (t_m, t_{m+1}]$. Since $\beta(t_m) = 0$ for all $m \in \mathbb{Z}$, there exists $\tilde{t} \in (t_m, t_{m+1})$ such that $\cos(\phi(\tilde{t}) + \beta(\tilde{t})) > \cos(\phi(\tilde{t}))$. However, by the assumption and the above derivatives, we know that

$$1 > \frac{a(\tilde{t})}{a(\tilde{t}) + \alpha(\tilde{t})} = \frac{\cos(\phi(\tilde{t}) + \beta(\tilde{t}))}{\cos(\phi(\tilde{t}))}, \quad (\text{A.10})$$

which is absurd. Thus, we have obtained the first claim.

The second claim could be obtained by taking Lemma 4 into account. Indeed, since $\beta'(t_n) = \frac{-\phi'(t_n)}{a(t_n) + \alpha(t_n)}\alpha(t_n)$ and $|\alpha(t)| \leq 2\pi\epsilon$, when ϵ is small enough, $|\beta'(t_n)| \leq \frac{2\phi'(t_n)}{a(t_n)}|\alpha(t_n)| \leq \frac{4\pi\phi'(t_n)}{a(t_n)}$. \square

Thus we obtain the control of the amplitude. Note that the proof does not depend on the condition about β'' .

Lemma 7. $|\beta''(t)| \leq 2\pi\epsilon$, $|\beta'(t)| \leq \frac{2\pi\epsilon}{c_1}$ and $|\beta(t)| \leq \frac{2\pi\epsilon}{c_1^2}$ for all $t \in \mathbb{R}$.

Proof. Suppose there existed $t' \in (t_m, t_{m+1})$ for some $m \in \mathbb{Z}$ so that $|\beta''(t')| > 3\pi\epsilon$. Without loss of generality, we assume $\beta''(t') > 0$. From (A.4) and (A.7) we have

$$|\beta'''(t)| \leq \epsilon(2\phi'(t) + \beta'(t)).$$

Thus, by the fundamental theorem of calculus, for any $t \in (t_m, t')$, we know

$$|\beta''(t') - \beta''(t)| \leq \int_t^{t'} |\beta'''(u)| du \leq \epsilon \int_t^{t'} (2\phi'(u) + \beta'(u)) du \leq 2\pi\epsilon,$$

where the last inequality holds due to Lemma 2 and the fact that $\phi(t') - \phi(t) \leq \phi(t_{m+1}) - \phi(t_m) = \pi$ and $|\beta(t') - \beta(t)| < \pi$ from Lemma 2. Similarly, we have that for all $t \in (t', t_{m+1})$, $|\beta''(t') - \beta''(t)| \leq 2\pi\epsilon$. Thus, $\beta''(t) > 0$ for all $t \in [t_m, t_{m+1}]$, which contradicts the fact that $\beta''(t)$ must change sign inside $[t_m, t_{m+1}]$ by Lemma 5.

With the upper bound of $|\beta''|$, we immediately have for all $t \in [t_m, t_{m+1}]$ that

$$|\beta'(t) - \beta'(t_m)| \leq \int_{t_m}^t |\beta''(u)| du \leq 2\pi(t - t_m)\epsilon.$$

To bound the right hand side, note that $t - t_m \leq t_{m+1} - t_m \leq \frac{\pi}{\phi'(t')}$, where $t' \in [t_m, t_{m+1}]$. Since $|\beta''(t)| \leq 2\pi\epsilon$, when ϵ is small enough, $\frac{\pi}{\phi'(t')} \leq \frac{2\pi}{\phi'(t)}$. Thus, $|\beta'(t) - \beta'(t_m)| \leq 2\pi(t - t_m)\epsilon \leq \frac{4\pi^2}{\phi'(t)}\epsilon$. To finish the proof, note that by Lemma 4, $|\beta'(t_m)| \leq \frac{4\pi\phi'(t_n)}{a(t_n)}\epsilon$. Similarly, we have the bound for β . \square

Appendix B. Proof of Theorem 2.2

When there are more than one gIMT in a given oscillatory signal $f \in \mathcal{Q}_{\epsilon, d}^{c_1, c_2, c_3}$, we lose the control of the hinging points for each gIMT like those, t_m and s_m , in Theorem 2.1. So the proof will be more qualitative. Suppose $f = \tilde{f} \in \mathcal{Q}_{\epsilon, d}^{c_1, c_2, c_3}$, where

$$f(t) = \sum_{l=1}^N a_l(t) \cos[2\pi\phi_l(t)], \quad \tilde{f}(t) = \sum_{l=1}^M A_l(t) \cos[2\pi\varphi_l(t)].$$

Fix $t_0 \in \mathbb{R}$. Denote $f_{t_0} := \sum_{l=1}^N f_{t_0, l}$, $\tilde{f}_{t_0} := \sum_{l=1}^M \tilde{f}_{t_0, l}$,

$$f_{t_0, l}(t) := a_l(t_0) \cos \left[2\pi \left(\phi_l(t_0) + \phi'_l(t_0)(t - t_0) + \phi''_l(t_0) \frac{(t - t_0)^2}{2} \right) \right]$$

and

$$\tilde{f}_{t_0, l}(t) := A_l(t_0) \cos \left[2\pi \left(\varphi_l(t_0) + \varphi'_l(t_0)(t - t_0) + \varphi''_l(t_0) \frac{(t - t_0)^2}{2} \right) \right].$$

Note that $f_{t_0, l}$ is an approximation of $a_l(t) \cos[2\pi\phi_l(t)]$ near t_0 based on the assumption of $\mathcal{Q}_{\epsilon, d}^{c_1, c_2, c_3}$, where we approximate the amplitude $a_l(t)$ by the zero-th order Taylor expansion and the phase function $\phi_l(t)$ by the second order Taylor expansion. To simplify the proof, we focus on the case that $|\phi''_l(t_0)| > \epsilon|\phi'_l(t_0)|$ and $|\varphi''_l(t_0)| > \epsilon|\varphi'_l(t_0)|$ for all l . For the case when there is one or more l so that $|\phi''_l(t_0)| \leq \epsilon|\phi'_l(t_0)|$, the proof follows the same line while we approximate the phases of these oscillatory components by the first order Taylor expansion.

Recall that the short time Fourier transform (STFT) of a given tempered distribution $f \in \mathcal{S}'$ associated with a Schwartz function $g \in \mathcal{S}$ as the window function is defined as

$$V_f^{(g)}(t, \eta) := \int_{\mathbb{R}} f(x) g(x - t) e^{-i2\pi\eta x} dx.$$

Note that by definition $f, f_{t_0}, \tilde{f}_{t_0} \in \mathcal{S}'$. To prove the theorem, we need the following lemma about the STFT.

Lemma 8. *For a fixed $t_0 \in \mathbb{R}$, we have*

$$\left| V_f^{(g)}(\tau, \eta) - V_{f_{t_0}}^{(g)}(\tau, \eta) \right| = O(\epsilon).$$

where C is a universal constant depending on c_1, c_2 and d .

Proof. Fix a time $t_0 \in \mathbb{R}$. By the same argument as that in [15, 9] and the conditions of

$\mathcal{Q}_{\epsilon, d}^{c_1, c_2, c_3}$, we immediately have

$$\begin{aligned}
& |V_f^{(g)}(\tau, \eta) - V_{f_{t_0}}^{(g)}(\tau, \eta)| \\
&= \left| \int_{\mathbb{R}} (f(t) - f_{t_0}(t))g(t - \tau)e^{-i2\pi\eta t} dt \right| \\
&\leq \sum_{l=1}^N \left| \int_{\mathbb{R}} (a_l(t) - a_l(t_0)) \cos[2\pi\phi_l(t)]g(t - t_0)e^{-i2\pi\eta t} dt \right| \\
&\quad + \sum_{l=1}^N \left| \int_{\mathbb{R}} a_l(t_0) (\cos[2\pi\phi_l(t)] - \cos[2\pi(\phi(t_0) + \phi'(t_0)(t - t_0) \right. \\
&\quad \left. + \frac{1}{2}\phi''(t_0)(t - t_0)^2)])g(t - t_0)e^{-i2\pi\eta t} dt \right| \\
&= O(\epsilon),
\end{aligned}$$

where the last term depends only on the first few absolute moments of g and g' , d , c_1 and c_2 . \square

With this claim, we know in particular that $V_f^{(g)}(t_0, \eta) = V_{f_{t_0}}^{(g)}(t_0, \eta) + O(\epsilon)$. As a result, the spectrogram of f and f_{t_0} are related by

$$|V_f^{(g)}(t_0, \eta)|^2 = |V_{f_{t_0}}^{(g)}(t_0, \eta)|^2 + O(\epsilon).$$

Next, recall that the spectrogram of a signal is intimately related to the Wigner-Ville distribution in the following way

$$|V_{f_{t_0}}^{(g)}(\tau, \eta)|^2 = \int \int W V_{f_{t_0}}(x, \xi) W V_g(x - \tau, \xi - \eta) dx d\xi,$$

where the Wigner-Ville distribution of a function h in the suitable space is defined as

$$W V_h(x, \xi) := \int h(x + \tau/2) h^*(x - \tau/2) e^{-i2\pi\tau\xi} d\tau.$$

Lemma 9. Take $g(t) = (2\sigma)^{1/4} \exp\{-\pi\sigma t^2\}$, where $\sigma = c_3$. When d is large enough described in (B.3), we have

$$|V_f^{(g)}(t_0, \eta)|^2 = L(t_0, \eta) + \epsilon \quad \text{and} \quad |V_{\tilde{f}}^{(g)}(t_0, \eta)|^2 = \tilde{L}(t_0, \eta) + \epsilon,$$

where

$$L(t_0, \eta) := \sum_{l=1}^N a_l^2(t_0) \sqrt{\frac{\sigma}{2(\sigma^2 + \phi_l''(t_0)^2)}} \exp\left\{-\frac{2\pi\sigma(\phi_l'(t_0) - \eta)^2}{\sigma^2 + \phi_l''(t_0)^2}\right\}$$

and

$$\tilde{L}(t_0, \eta) := \sum_{l=1}^M A_l^2(t_0) \sqrt{\frac{\sigma}{2(\sigma^2 + \varphi_l''(t_0)^2)}} \exp\left\{-\frac{2\pi\sigma(\varphi_l'(t_0) - \eta)^2}{\sigma^2 + \varphi_l''(t_0)^2}\right\}.$$

Proof. By a direct calculation, the Wigner-Ville distribution of the Gaussian function $g(t) = (2\sigma)^{1/4} \exp\{-\pi\sigma t^2\}$ with the unit energy, where $\sigma > 0$, is

$$WV_g(x, \xi) = 2 \exp\left\{-2\pi\left(\sigma x^2 + \frac{\xi^2}{\sigma}\right)\right\};$$

similarly, the Wigner-Ville distribution of $f_{t_0, l}$ is

$$WV_{f_{t_0, l}}(x, \xi) = a_l^2(t_0) \delta_{\phi'_l(t_0) + \phi''_l(t_0)(x-t_0)}(\xi).$$

Thus, we know

$$\begin{aligned} & \left| V_{f_{t_0, l}}^{(g)}(t_0, \eta) \right|^2 \\ &= \int \int WV_{f_{t_0, l}}(x, \xi) WV_g(x - t_0, \xi - \eta) dx d\xi \\ &= \int \int \left(a_l^2(t_0) \delta_{\phi'_l(t_0) + \phi''_l(t_0)(x-t_0)}(\xi) \right) 2 \exp\left\{-2\pi\left(\sigma(x-t_0)^2 + \frac{(\xi-\eta)^2}{\sigma}\right)\right\} d\xi dx \\ &= 2a_l^2(t_0) \int \exp\left\{-2\pi\left(\sigma(x-t_0)^2 + \frac{(\phi'_l(t_0) + \phi''_l(t_0)(x-t_0) - \eta)^2}{\sigma}\right)\right\} dx \\ &= a_l^2(t_0) \sqrt{\frac{\sigma}{2(\sigma^2 + \phi''_l(t_0)^2)}} \exp\left\{-\frac{2\pi\sigma}{\sigma^2 + \phi''_l(t_0)^2}(\phi'_l(t_0) - \eta)^2\right\}. \end{aligned} \tag{B.1}$$

Thus, we have the expansion of $\sum_{l=1}^N |V_{f_{t_0, l}}^{(g)}(t_0, \eta)|^2$, which is $L(t_0, \eta)$. Next, we clearly have

$$\begin{aligned} & \left| V_{f_{t_0}}^{(g)}(\tau, \eta) \right|^2 - \sum_{l=1}^N |V_{f_{t_0, l}}^{(g)}(\tau, \eta)|^2 = \left| \Re \sum_{k \neq l} V_{f_{t_0, l}}^{(g)}(\tau, \eta) \overline{V_{f_{t_0, k}}^{(g)}(\tau, \eta)} \right| \\ & \leq \sum_{k \neq l} \left| V_{f_{t_0, l}}^{(g)}(\tau, \eta) \right| \left| V_{f_{t_0, k}}^{(g)}(\tau, \eta) \right|. \end{aligned}$$

To bound the right hand side, note that (B.1) implies

$$\left| V_{f_{t_0, l}}^{(g)}(t_0, \eta) \right| = a_l(t_0) \left(\frac{\sigma}{2(\sigma^2 + \phi''_l(t_0)^2)} \right)^{1/4} \exp\left\{-\frac{\pi\sigma(\phi'_l(t_0) - \eta)^2}{\sigma^2 + \phi''_l(t_0)^2}\right\}.$$

As a result, $\sum_{k \neq l} \left| V_{f_{t_0, l}}^{(g)}(t_0, \eta) \right| \left| V_{f_{t_0, k}}^{(g)}(t_0, \eta) \right|$ becomes

$$\sum_{k \neq l} \frac{a_k(t_0) a_l(t_0) \sigma^{1/2}}{(4(\sigma^2 + \phi''_k(t_0)^2)(\sigma^2 + \phi''_l(t_0)^2))^{1/4}} \exp\left\{-\pi\sigma\left(\frac{(\phi'_k(t_0) - \eta)^2}{\sigma^2 + \phi''_k(t_0)^2} + \frac{(\phi'_l(t_0) - \eta)^2}{\sigma^2 + \phi''_l(t_0)^2}\right)\right\},$$

which is a smooth and bounded function of η and is bounded by

$$\frac{c_2^2}{\sqrt{2}\sigma^{1/2}} \sum_{k \neq l} \exp\left\{-\frac{\pi\sigma}{\sigma^2 + c_3^2} [(\phi'_k(t_0) - \eta)^2 + (\phi'_l(t_0) - \eta)^2]\right\}. \tag{B.2}$$

Suppose the maximum of the right hand side is achieved when $\eta = \phi'_{k_0}(t_0)$, where $k_0 = \arg \min_{l=1, \dots, N-1} (\phi'_{l+1}(t_0) - \phi'_l(t_0))$. To bound (B.2), denote $\gamma = \frac{\pi\sigma}{\sigma^2 + c_3^2}$ to simplify the notation. Clearly, the summation in (B.2) is thus bounded by

$$2 \sum_{l=1}^{\infty} e^{-l^2 \gamma d^2} + 2 \sum_{k=1}^{\infty} e^{-k^2 \gamma d^2} \sum_{l=0}^{\infty} e^{-l^2 \gamma d^2} \leq 2(Q+1)S,$$

where $Q = \int_0^{\infty} e^{-l^2 \gamma d^2 t} dt = \frac{\sqrt{\pi}}{2\gamma d^2}$ and $S = \int_1^{\infty} e^{-l^2 \gamma d^2 t} dt \leq \frac{d\sqrt{\gamma}}{e^{-\gamma d^2}}$. Note that here we take the bounds $\sum_{l=0}^{\infty} e^{-l^2 \gamma d^2} \leq Q$ and $\sum_{l=1}^{\infty} e^{-l^2 \gamma d^2} \leq S$. Thus, we conclude that the interference term is bounded by $\frac{\sqrt{2}c_3^2}{\sigma^{1/2}}(Q+1)S$. To finish the proof, we require the interference term to be bounded by ϵ , which leads to the following bound of d when we take $\sigma = c_3$:

$$d \geq \sqrt{2 \ln c_2 + \frac{1}{2} \ln c_3 - \ln \epsilon}. \quad (\text{B.3})$$

We have finished the proof. \square

Since t_0 is arbitrary in the above argument and the spectrogram of a function is unique, we have $|V_f^{(g)}(t_0, \eta)|^2 = |V_{\tilde{f}}^{(g)}(t_0, \eta)|^2$ and hence

$$|L(t_0, \eta) - \tilde{L}(t_0, \eta)| = O(\epsilon). \quad (\text{B.4})$$

With the above claim, we now show $M = N$.

Lemma 10. $M = N$.

Proof. With $\sigma = c_3$, by Lemma 9, for each $l = 1, \dots, N$, there exists a subinterval $I_l(t_0)$ around $\phi'_l(t_0)$ so that on $I_l(t_0)$, $L(t_0, \eta) > \frac{a_l^2(t_0)c_3^2}{2\sqrt{2(c_3^2 + \phi_l''(t_0)^2)}} > \frac{c_1^2 c_3^2}{2\sqrt{2c_3^2 + 2c_2^2}}$. Similarly, for each $l = 1, \dots, M$, there exists a subinterval $J_l(t_0)$ around $\varphi'_l(t_0)$ so that on $J_l(t_0)$, $\tilde{L}(t_0, \eta) > \frac{A_l^2(t_0)c_3^2}{2\sqrt{2(c_3^2 + \varphi_l''(t_0)^2)}} > \frac{c_1^2 c_3^2}{2\sqrt{2c_3^2 + 2c_2^2}}$. Thus, when ϵ is small enough, in particular, $\epsilon \ll \frac{c_1^2 c_3^2}{2\sqrt{2c_3^2 + 2c_2^2}}$, the equality in (B.4) cannot hold if $M \neq N$. \square

With this claim, we obtain the first part of the proof, and hence the equality

$$f(t) = \sum_{l=1}^N a_l(t) \cos[2\pi\phi_l(t)] = \sum_{l=1}^N A_l(t) \cos[2\pi\varphi_l(t)] \in \mathcal{Q}_{\epsilon, d}^{c_1, c_2, c_3}. \quad (\text{B.5})$$

Now we proceed to finish the proof. Note that it is also clear that the sets $I_l(t_0)$ and $J_l(t_0)$ defined in the proof of Lemma 10 satisfy that $I_l(t_0) \cap I_k(t_0) = \emptyset$ for all $l \neq k$. Also, $I_l(t_0) \cap J_l(t_0) \neq \emptyset$ and $I_l(t_0) \cap J_k(t_0) = \emptyset$ for all $l \neq k$. Indeed, if $k = l + 1$ and we have $I_l(t_0) \cap J_{l+1}(t_0) \neq \emptyset$, then $L(t_0, \eta) > \frac{c_1^2 c_3^2}{2\sqrt{2c_3^2 + 2c_2^2}}$ on $J_{l+1}(t_0) \setminus I_l(t_0)$, which leads to the contradiction. By the ordering of $\phi'_l(t_0)$ and hence the ordering of $I_l(t_0)$, we have the result.

Take $\ell = 1$ and $\eta = \phi'_1(t_0)$. By Lemma 9, when d is large enough, on $I_1(t_0)$ we have

$$\frac{a_1^2(t_0)}{\sqrt{2(1 + \phi_1''(t_0)^2)}} = \frac{A_1^2(t_0)}{\sqrt{2(1 + \varphi_1''(t_0)^2)}} \exp \left\{ -\frac{2\pi c_3(\varphi_1'(t_0) - \phi_1'(t_0))^2}{c_3^2 + \varphi_1''(t_0)^2} \right\} + O(\epsilon), \quad (\text{B.6})$$

which leads to the fact that

$$\left| \frac{a_1^2(t_0)c_3^2}{\sqrt{2(c_3^2 + \phi_1''(t_0)^2)}} - \frac{A_1^2(t_0)c_3^2}{\sqrt{2(c_3^2 + \varphi_1''(t_0)^2)}} \right| = O(\epsilon). \quad (\text{B.7})$$

Indeed, without loss of generality, assume $\frac{a_1^2(t_0)c_3^2}{\sqrt{2(c_3^2 + \phi_1''(t_0)^2)}} \geq \frac{A_1^2(t_0)c_3^2}{\sqrt{2(c_3^2 + \varphi_1''(t_0)^2)}}$ and we have

$$\begin{aligned} & \frac{a_1^2(t_0)c_3^2}{\sqrt{2(c_3^2 + \phi_1''(t_0)^2)}} - \frac{A_1^2(t_0)c_3^2}{\sqrt{2(c_3^2 + \varphi_1''(t_0)^2)}} \\ & \leq \frac{a_1^2(t_0)c_3^2}{\sqrt{2(c_3^2 + \phi_1''(t_0)^2)}} - \frac{A_1^2(t_0)c_3^2}{\sqrt{2(c_3^2 + \varphi_1''(t_0)^2)}} \exp \left\{ -\frac{2\pi c_3(\varphi_1'(t_0) - \phi_1'(t_0))^2}{c_3^2 + \varphi_1''(t_0)^2} \right\} = O(\epsilon) \end{aligned}$$

by (B.6) since 0 is the unique maximal point of the chosen Gaussian function.

Lemma 11. $|\phi'_\ell(t) - \varphi'_\ell(t)| = O(\sqrt{\epsilon})$ for all time $t \in \mathbb{R}$ and $\ell = 1, \dots, N$.

Proof. Fix $t_0 \in \mathbb{R}$ and $\ell = 1$. By (B.6), (B.7) and the conditions of $\mathcal{Q}_{\epsilon, d}^{c_1, c_2, c_3}$, on $I_1(t_0)$ we have

$$\left| \frac{A_1^2(t_0)c_3^2}{\sqrt{2(c_3^2 + \varphi_1''(t_0)^2)}} \left| 1 - \exp \left\{ -\frac{2\pi c_3(\varphi_1'(t_0) - \phi_1'(t_0))^2}{c_3^2 + \varphi_1''(t_0)^2} \right\} \right| \right| = O(\epsilon).$$

Due to the fact that the Gaussian function monotonically decreases as $\frac{2\pi c_3(\varphi_1'(t_0) - \phi_1'(t_0))^2}{c_3^2 + \varphi_1''(t_0)^2} > 0$, we have

$$\frac{(\varphi_1'(t_0) - \phi_1'(t_0))^2}{c_3^2 + \varphi_1''(t_0)^2} = O(\epsilon).$$

Since φ_1'' is uniformly bounded by c_2 , we know

$$|\varphi_1'(t_0) - \phi_1'(t_0)| = O(\sqrt{\epsilon}).$$

By the same argument, we know that $|\varphi'_l(t) - \phi'_l(t)| = O(\sqrt{\epsilon})$ for all $l = 1, \dots, N$ and $t \in \mathbb{R}$. \square

Lemma 12. $|\phi''_\ell(t) - \varphi''_\ell(t)| = O(\sqrt{\epsilon})$ for all time $t \in \mathbb{R}$ and $\ell = 1, \dots, N$.

Proof. Fix $t_0 \in \mathbb{R}$ and $\ell = 1$. By the assumption that $\phi_1'''(t_0) = O(\epsilon)$ and $\varphi_1'''(t_0) = O(\epsilon)$, we claim that $|\phi_1''(t_0) - \varphi_1''(t_0)| = O(\sqrt{\epsilon})$ holds. Indeed, we have

$$\phi_1'(t_0 + 1) = \phi_1'(t_0) + \int_{t_0}^{t_0+1} \phi_1''(s) ds \quad \text{and} \quad \varphi_1'(t_0 + 1) = \varphi_1'(t_0) + \int_{t_0}^{t_0+1} \varphi_1''(s) ds,$$

which leads to the relationship

$$\phi_1'(t_0 + 1) - \varphi_1'(t_0 + 1) = \phi_1'(t_0) - \varphi_1'(t_0) + \int_{t_0}^{t_0+1} (\phi_1''(s) - \varphi_1''(s)) ds.$$

Therefore, by the assumption that $\phi_1'''(t_0) = O(\epsilon)$ and $\varphi_1'''(t_0) = O(\epsilon)$, we have

$$\begin{aligned} & \int_{t_0}^{t_0+1} (\phi_1''(s) - \varphi_1''(s)) ds \\ &= \int_{t_0}^{t_0+1} \left(\phi_1''(t_0) - \varphi_1''(t_0) + \int_{t_0}^s (\phi_1'''(x) - \varphi_1'''(x)) dx \right) ds \\ &= \phi_1''(t_0) - \varphi_1''(t_0) + O(\epsilon), \end{aligned}$$

which means that $|\phi_1''(t_0) - \varphi_1''(t_0)| = O(\sqrt{\epsilon})$ since $|\phi_1'(t_0+1) - \varphi_1'(t_0+1)| = O(\sqrt{\epsilon})$ and $|\phi_1'(t_0) - \varphi_1'(t_0)| = O(\sqrt{\epsilon})$. By the same argument, we know that $|\varphi_l''(t) - \phi_l''(t)| = O(\sqrt{\epsilon})$ for all $l = 1, \dots, N$ and $t \in \mathbb{R}$. \square

Lemma 13. $|a_\ell(t) - A_\ell(t)| = O(\sqrt{\epsilon})$ for all time $t \in \mathbb{R}$ and $\ell = 1, \dots, N$.

Proof. Fix $t_0 \in \mathbb{R}$ and $\ell = 1$. From (B.7), it is clear that $|a_1(t_0) - A_1(t_0)| = O(\sqrt{\epsilon})$ if and only if $|\phi_1''(t_0) - \varphi_1''(t_0)| = O(\sqrt{\epsilon})$, so we obtain the claim by Lemma 12. Similar argument holds for all time $t \in \mathbb{R}$ and $\ell = 2, \dots, N$. \square

Lastly, we show the difference of the phase functions.

Lemma 14. $|\phi_\ell(t) - \varphi_\ell(t)| = O(\sqrt{\epsilon})$ for all time $t \in \mathbb{R}$ and $\ell = 1, \dots, N$.

Proof. By (B.5) and the fact that $|a_l(t) - A_l(t)| = O(\sqrt{\epsilon})$, we have for all $t \in \mathbb{R}$,

$$\sum_{l=1}^N a_l(t) \cos[2\pi\phi_l(t)] = \sum_{l=1}^N a_l(t) \cos[2\pi(\phi_l(t) + \alpha_l(t))] + O(\sqrt{\epsilon}),$$

where $\alpha_l \in C^3(\mathbb{R})$. Note that $\sum_{l=1}^N a_l(t) \cos[2\pi(\phi_l(t) + \alpha_l(t))] \in \mathcal{Q}_{\epsilon,d}^{c_1, c_2, c_3}$. Fix $t_0 \in \mathbb{R}$. Suppose there exists t_0 and the smallest number k so that $\alpha_k(t_0) = O(\sqrt{\epsilon})$ up to multiples of 2π does not hold. Then there exists at least one $\ell \neq k$ so that $\alpha_\ell(t_0) = O(\sqrt{\epsilon})$ does not hold. Suppose $L > k$ is the largest integer that $\alpha_L(t_0) = O(\sqrt{\epsilon})$ does not hold. In this case, there exists $t_1 > t_0$ so that $\sum_{l=1}^N a_l(t_1) \cos[2\pi\phi_l(t_1)] = \sum_{l=1}^N a_l(t_1) \cos[2\pi(\phi_l(t_1) + \alpha_l(t_1))] + O(\sqrt{\epsilon})$ does not hold. Indeed, as $\phi_L'(t_0)$ is higher than $\phi_k'(t_0)$ by at least d , we could find $t_1 = \phi_k^{-1}(\phi_k(t_0) + c)$, where $0 < c < \pi$, so that $\cos[2\pi(\phi_L(t_1) + \alpha_L(t_1))] - \cos[2\pi(\phi_L(t_1))] = \cos[2\pi(\phi_L(t_0) + \alpha_L(t_0))] - \cos[2\pi(\phi_L(t_0))] + O(\sqrt{\epsilon})$ does not hold while $\sum_{l \neq L}^N a_l(t) \cos[2\pi\phi_l(t)] = \sum_{l \neq L}^N a_l(t) \cos[2\pi(\phi_l(t) + \alpha_l(t))] + O(\sqrt{\epsilon})$ holds. We thus get a contradiction and hence the proof. \square

Appendix C. A convergence study of Algorithm 2

We provide here a simple convergence study of Algorithm 2, based on the Zangwill's global convergence theorem [53] which can be stated as follow

Theorem Appendix C.1. *Let \mathcal{A} be an algorithm on \mathcal{X} , and suppose that, given $x_0 \in \mathcal{X}$, the sequence $\{x_k\}_{k=1}^\infty$ is generated and satisfies*

$$x_{k+1} \in \mathcal{A}(x_k) \ .$$

Let a solution set Γ be given, and suppose that

- (i) the sequence $\{x_k\}_{k=0}^\infty \subset \mathcal{S}$ for $\mathcal{S} \subset \mathcal{X}$ a compact set.
- (ii) there is a continuous function \mathcal{Z} on \mathcal{X} such that
 - (a) if $x \notin \Gamma$, then $\mathcal{Z}(y) < \mathcal{Z}(x)$ for all $y \in \mathcal{A}(x)$
 - (b) if $x \in \Gamma$, then $\mathcal{Z}(y) \leq \mathcal{Z}(x)$ for all $y \in \mathcal{A}(x)$
- (iii) the mapping \mathcal{A} is closed at all point $\mathcal{X} \setminus \Gamma$

Then the limit of any convergent subsequence of $\{x_k\}_{k=0}^\infty$ is a solution, and $\mathcal{Z}_k \rightarrow \mathcal{Z}(x^*)$ for some $x^* \in \Gamma$.

The algorithm \mathcal{A} is here the alternating minimization Alg. 2. The solution set Γ is then naturally the set of the critical points of the functional \mathcal{H} . Equivalently, Γ is the set of the fixed point of Alg. 2. Indeed, for α being fixed, \mathbf{F}^* is a minimizer of \mathcal{H}_α is equivalent to \mathbf{F}^* is a fixed point of Alg. 1 [14]. The descent function \mathcal{Z} is then naturally the functional \mathcal{H} .

As we work in the finite dimensional case, the boundeness of the sequence and then point (i) of the Theorem is here direct consequences of point (iii). Moreover, thanks to the continuity of the soft-thresholding operator, the mapping \mathcal{A} is continuous and point (iii) is also a direct consequence of point (iii).

Point (iii) of the theorem comes from the monotonic version of FISTA. Indeed, the very first iteration of FISTA is equivalent to a "simple" forward-backward step, which ensure the strict decreasing of the functional \mathcal{H}_α (see [48]). Then, as α is the unique minimizer of the function $\mathcal{H}_\mathbf{F}$, we have a sequence $\{(\alpha_k, \mathbf{F}_k)\}_{k=0}^\infty$ such that $\mathcal{H}(\alpha^{k+1}, \mathbf{F}^{k+1}) < \mathcal{H}(\alpha^k, \mathbf{F}^k)$ as soon as (α^k, \mathbf{F}^k) is not a critical point of \mathcal{H} .